

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : A61K 39/02		A1	(11) International Publication Number: WO 94/21290 (43) International Publication Date: 29 September 1994 (29.09.94)
(21) International Application Number: PCT/US94/02550 (22) International Filing Date: 15 March 1994 (15.03.94)		(81) Designated States: AU, BR, CA, FI, JP, KR, NO, RU, UA, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).	
(30) Priority Data: 038,682 16 March 1993 (16.03.93) US		Published <i>With international search report.</i>	
(71)(72) Applicants and Inventors: BARENKAMP, Stephen, J. [US/US]; 16 Villawood Lane, Webster Grove, MO 63119-4954 (US). ST. GEME, Joseph, William, III [US/US]; 45 Bershire Drive, St. Louis, MO 63117 (US).			
(74) Agent: BERKSTRESSER, Jerry, W.; Shoemaker and Mattare, Ltd., 2001 Jefferson Davis Highway, 1203 Crystal Plaza Building 1, P.O. Box 2286, Arlington, VA 22202-0286 (US).			
(54) Title: HIGH MOLECULAR WEIGHT SURFACE PROTEINS OF NON-TYPEABLE HAEMOPHILUS			
(57) Abstract <p>High molecular weight surface proteins of non-typeable <i>Haemophilus influenzae</i> which exhibit immunogenic properties and genes encoding the same are described. Specifically, genes coding for two immunodominant high molecular weight proteins, HMW1 and HMW2, have been cloned, expressed and sequenced, while genes coding for high molecular proteins HMW3 and HMW4 have been cloned, expressed and partially sequenced.</p>			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CS	Czechoslovakia	LV	Latvia	TG	Togo
CZ	Czech Republic	MC	Monaco	TJ	Tajikistan
DE	Germany	MD	Republic of Moldova	TT	Trinidad and Tobago
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	US	United States of America
FI	Finland	MN	Mongolia	UZ	Uzbekistan
FR	France			VN	Viet Nam
GA	Gabon				

TITLE OF INVENTIONHIGH MOLECULAR WEIGHT SURFACE PROTEINS
OF NON-TYPEABLE HAEMOPHILUSFIELD OF INVENTION

5 This invention relates to high molecular weight proteins of non-typeable haemophilus.

BACKGROUND TO THE INVENTION

10 Non-typeable Haemophilus influenzae are non-encapsulated organisms that are defined by their lack of reactivity with antisera against known H. influenzae capsular antigens.

15 These organisms commonly inhabit the upper respiratory tract of humans and are frequently responsible for infections, such as otitis media, sinusitis, conjunctivitis, bronchitis and pneumonia. Since these organisms do not have a polysaccharide capsule, they are not controlled by the present Haemophilus influenzae type b (Hib) vaccines, which are directed towards Hib bacterial capsular polysaccharides.

20 The non-typeable strains, however, do produce surface antigens that can elicit bactericidal antibodies. Two of the major outer membrane proteins, P2 and P6, have been identified as targets of human serum bactericidal activity. However, it has been shown that the P2 protein sequence is variable, in particular in the non-typeable Haemophilus strains. Thus, a P2-based vaccine would not protect against all strains of the organism.

25 There have previously been identified by Barenkamp et al (Pediatr. Infect. Dis. J., 9:333-339, 1990) a group of high-molecular-weight (HMW) proteins that appeared to be major targets of antibodies present in human convalescent sera. Examination of a series of middle ear isolates revealed the presence of one or two such proteins in most strains. However, prior to the present invention, the structures of these proteins were unknown as were pure isolates of such proteins.

SUBSTITUTE SHEET (RULE 26)

SUMMARY OF INVENTION

The inventors, in an effort to further characterize the high molecular weight (HMW) Haemophilus proteins, have cloned, expressed and sequenced the genes coding for two immunodominant HMW proteins (designated HMW1 and HMW2) from a prototype non-typeable Haemophilus strain and have cloned, expressed and almost completely sequenced the genes coding for two additional immunodominant HMW proteins (designated HMW3 and HMW4) 5 from another non-typeable Haemophilus strain. 10

In accordance with one aspect of the present invention, therefore, there is provided an isolated and purified gene coding for a high molecular weight protein 15 of a non-typeable Haemophilus strain, particularly a gene coding for protein HMW1, HMW2, HMW3 or HMW4, as well as any variant or fragment of such protein which retains the immunological ability to protect against disease caused by a non-typeable Haemophilus strain. In another aspect, the invention provides a high molecular weight protein of 20 non-typeable Haemophilus influenzae which is encoded by these genes.

BRIEF DESCRIPTION OF DRAWINGS

Figure 1 is a DNA sequence of a gene coding for protein HMW1 (SEQ ID NO: 1);

25 Figure 2 is a derived amino acid sequence of protein HMW1 (SEQ ID NO: 2);

Figure 3 is a DNA sequence of a gene coding for protein HMW2 (SEQ ID NO: 3);

30 Figure 4 is a derived amino acid sequence of HMW2 (SEQ ID NO: 4);

Figure 5A shows restriction maps of representative recombinant phages which contained the HMW1 or HMW2 structural genes, the locations of the structural genes being indicated by the shaded bars;

35 Figure 5B shows the restriction map of the T7 expression vector pT7-7;

Figure 6 contains the DNA sequence of a gene cluster for the hmw1 gene (SEQ ID NO: 5), comprising nucleotides 351 to 4958 (ORF a) (as in Figure 1), as well as two additional downstream genes in the 3' flanking region, comprising ORFs b, nucleotides 5114-6748 and c nucleotides 7062-9011;

Figure 7 contains the DNA sequence of a gene cluster for the hmw2 gene (SEQ ID NO: 6), comprising nucleotides 792 to 5222 (ORF a) (as in Figure 3), as well as two additional downstream genes in the 3' flanking region, comprising ORFs b, nucleotides 5375-7009, and c, nucleotides 7249-9198;

Figure 8 is a partial DNA sequence of a gene coding for protein HMW3 (SEQ ID NO: 7);

Figure 9 is a partial DNA sequence of a gene coding for protein HMW4 (SEQ ID NO: 8); and

Figure 10 is a comparison table for the derived amino acid sequence for proteins HMW1, HMW2, HMW3 and HMW4.

GENERAL DESCRIPTION OF INVENTION

The DNA sequences of the genes coding for HMW1 and HMW2, shown in Figures 1 and 3 respectively, were shown to be about 80% identical, with the first 1259 base pairs of the genes being identical. The derived amino acid sequences of the two HMW proteins, shown in Figures 2 and 4 respectively, are about 70% identical. Furthermore, the encoded proteins are antigenically related to the filamentous hemagglutinin surface protein of Bordetella pertussis. A monoclonal antibody prepared against filamentous hemagglutinin (FHA) of Bordetella pertussis was found to recognize both of the high molecular weight proteins. This data suggests that the HMW and FHA proteins may serve similar biological functions. The derived amino acid sequences of the HMW1 and HMW2 proteins show sequence similarity to that for the FHA protein. It has further been shown that these

5 antigenically-related proteins are produced by the majority of the non-typeable strains of Haemophilus. Antisera raised against the protein expressed by the HMW1 gene recognizes both the HMW2 protein and the B. pertussis FHA. The present invention includes an isolated and purified high molecular weight protein of non-typeable haemophilus which is antigenically related to the B. pertussis FHA, which may be obtained from natural sources or produced recombinantly.

10 10 A phage genomic library of a known strain of non-typeable Haemophilus was prepared by standard methods and the library was screened for clones expressing high molecular weight proteins, using a high titre antiserum against HMW's. A number of strongly reactive DNA clones were plaque-purified and sub-cloned into a T7 expression plasmid. It was found that they all expressed either one or the other of the two high-molecular-weight proteins designated HMW1 and HMW2, with apparent molecular weights of 125 and 120 kDa, respectively, encoded by open reading

15 15 frames of 4.6 kb and 4.4 kb, respectively.

20 20 Representative clones expressing either HMW1 or HMW2 were further characterized and the genes isolated, purified and sequenced. The DNA sequence of HMW1 is shown in Figure 1 and the corresponding derived amino acid sequence in Figure 2. Similarly, the DNA sequence of HMW2 is shown in Figure 3 and the corresponding derived amino acid sequence in Figure 4. Partial purification of the isolated proteins and N-terminal sequence analysis indicated that the expressed proteins are truncated since their sequence starts at residue number 442 of both full length HMW1 and HMW2 gene products.

25 25 Subcloning studies with respect to the hmw1 and hmw2 genes indicated that correct processing of the HMW proteins required the products of additional downstream genes. It has been found that both the hmw1 and hmw2 genes are flanked by two additional downstream open

reading frames (ORFs), designated b and c, respectively, (see Figures 6 and 7).

5 The b ORFs are 1635 bp in length, extending from nucleotides 5114 to 6748 in the case of hmw1 and nucleotides 5375 to 7009 in the case of hmw2, with their derived amino acid sequences 99% identical. The derived amino acid sequences demonstrate similarity with the derived amino acid sequences of two genes which encode proteins required for secretion and activation of 10 hemolysins of P. mirabilis and S. marcescens.

15 The c ORFs are 1950 bp in length, extending from nucleotides 7062 to 9011 in the case of hmw1 and nucleotides 7249 to 9198 in the case of hmw2, with their derived amino acid sequences 96% identical. The hmw1 c ORF is preceded by a series of 9 bp direct tandem repeats. In plasmid subclones, interruption of the hmw1 b or c ORF results in defective processing and secretion 20 of the hmw1 structural gene product.

25 The two high molecular weight proteins have been isolated and purified and shown to be partially protective against otitis media in chinchillas and to function as adhesins. These results indicate the potential for use of such high molecular proteins and structurally-related proteins of other non-typeable strains of Haemophilus influenzae as components in non-typeable Haemophilus influenzae vaccines.

30 Since the proteins provided herein are good cross-reactive antigens and are present in the majority of non-typeable Haemophilus strains, it is evident that these HMW proteins may become integral constituents of a universal Haemophilus vaccine. Indeed, these proteins may be used not only as protective antigens against otitis, sinusitis and bronchitis caused by the 35 non-typeable Haemophilus strains, but also may be used as carriers for the protective Hib polysaccharides in a conjugate vaccine against meningitis. The proteins also

may be used as carriers for other antigens, haptens and polysaccharides from other organisms, so as to induce immunity to such antigens, haptens and polysaccharides.

5 The nucleotide sequences encoding two high molecular weight proteins of a different non-typeable Haemophilus strain (designated HMW3 and HMW4) have been largely elucidated, and are presented in Figures 8 and 9. HMW3 has an apparent molecular weight of 125 kDa while HMW4 has an apparent molecular weight of 123 kDa. These high
10 molecular weight proteins are antigenically related to the HMW1 and HMW2 proteins and to FHA. Sequence analysis of HMW3 is approximately 85% complete and of HMW4 95% complete, with short stretches at the 5'-ends of each gene remaining to be sequenced.

15 Figure 10 contains a multiple sequence comparison of the derived amino acid sequences for the four high molecular weight proteins identified herein. As may be seen from this comparison, stretches of identical peptide sequence may be found throughout the length of the comparison, with HMW3 more closely resembling HMW1 and HMW4 more closely resembling HMW2. This information is highly suggestive of a considerable sequence homology between high molecular weight proteins from various non-typeable Haemophilus strains.

20 25 In addition, mutants of non-typeable H. influenzae strains that are deficient in expression of HMW1 or HMW2 or both have been constructed and examined for their capacity to adhere to cultured human epithelial cells. The hmw1 and hmw2 gene clusters have been expressed in E. coli and have been examined for in vitro adherence. The results of such experimentation demonstrate that both HMW1 and HMW2 mediate attachment and hence are adhesins and that this function is present even in the absence of other H. influenzae surface structures.

30 35 With the isolation and purification of the high molecular weight proteins, the inventors are able to

5 determine the major protective epitopes by conventional epitope mapping and synthesize peptides corresponding to these determinants to be incorporated in fully synthetic or recombinant vaccines. Accordingly, the invention also
10 comprises a synthetic peptide having an amino acid sequence corresponding to at least one protective epitope of a high molecular weight protein of a non-typeable Haemophilus influenzae. Such peptides are of varying length that constitute portions of the high-molecular-weight proteins, that can be used to induce immunity, either directly or as part of a conjugate, against the relative organisms and thus constitute vaccines for protection against the corresponding diseases.

15 The present invention also provides any variant or fragment of the proteins that retains the potential immunological ability to protect against disease caused by non-typeable Haemophilus strains. The variants may be constructed by partial deletions or mutations of the 20 genes and expression of the resulting modified genes to give the protein variations.

EXAMPLES

Example 1:

25 Non-typeable H.influenzae strains 5 and 12 were isolated in pure culture from the middle ear fluid of children with acute otitis media. Chromosomal DNA from strain 12, providing genes encoding proteins HMW1 and HMW2, was prepared by preparing Sau3A partial restriction 30 digests of chromosomal DNA and fractionating on sucrose gradients. Fractions containing DNA fragments in the 9 to 20 kbp range were pooled and a library was prepared by ligation into λ EMBL3 arms. Ligation mixtures were packaged in vitro and plate-amplified in a P2 lysogen of E. coli LE392.

35 For plasmid subcloning studies, DNA from a representative recombinant phage was subcloned into the

T7 expression plasmid pT7-7, containing the T7 RNA polymerase promoter Φ 10, a ribosome-binding site and the translational start site for the T7 gene 10 protein upstream from a multiple cloning site (see Figure 5B).

5 DNA sequence analysis was performed by the dideoxy method and both strands of the HMW1 gene and a single strand of the HMW2 gene were sequenced.

10 Western immunoblot analysis was performed to identify the recombinant proteins being produced by reactive phage clones. Phage lysates grown in LE392 cells or plaques picked directly from a lawn of LE392 cells on YT plates were solubilized in gel electrophoresis sample buffer prior to electrophoresis. Sodium dodecyl sulfate (SDS)-polyacrylamide gel 15 electrophoresis was performed on 7.5% or 11% polyacrylamide modified Laemmli gels. After transfer of the proteins to nitrocellulose sheets, the sheets were probed sequentially with an *E. coli*-absorbed human serum sample containing high-titer antibody to the high-molecular-weight proteins and then with alkaline phosphatase-conjugated goat anti-human immunoglobulin G (IgG) second antibody. Sera from healthy adults contains 20 high-titer antibody directed against surface-exposed high-molecular-weight proteins of non-typeable *H. influenzae*. One such serum sample was used as the screening antiserum after having been extensively 25 absorbed with LE392 cells.

30 To identify recombinant proteins being produced by *E. coli* transformed with recombinant plasmids, the plasmids of interest were used to transform *E. coli* BL21 (DE3)/pLysS. The transformed strains were grown to an A_{600} of 0.5 in L broth containing 50 μ g of ampicillin per ml. IPTG was then added to 1 mM. One hour later, cells 35 were harvested, and a sonicate of the cells was prepared. The protein concentrations of the samples were determined by the bicinchoninic acid method. Cell sonicates

containing 100 μ g of total protein were solubilized in electrophoresis sample buffer, subjected to SDS-polyacrylamide gel electrophoresis, and transferred to nitrocellulose. The nitrocellulose was then probed sequentially with the E. coli-absorbed adult serum sample and then with alkaline phosphatase-conjugated goat anti-human IgG second antibody.

Western immunoblot analysis also was performed to determine whether homologous and heterologous non-typeable H. influenzae strains expressed high-molecular-weight proteins antigenically related to the protein encoded by the cloned HMW1 gene (rHMW1). Cell sonicates of bacterial cells were solubilized in electrophoresis sample buffer, subjected to SDS-polyacrylamide gel electrophoresis, and transferred to nitrocellulose. Nitrocellulose was probed sequentially with polyclonal rabbit rHMW1 antiserum and then with alkaline phosphatase-conjugated goat anti-rabbit IgG second antibody.

Finally, Western immunoblot analysis was performed to determine whether non-typeable Haemophilus strains expressed proteins antigenically related to the filamentous hemagglutinin protein of Bordetella pertussis. Monoclonal antibody X3C, a murine immunoglobulin G (IgG) antibody which recognizes filamentous hemagglutinin, was used to probe cell sonicates by Western blot. An alkaline phosphatase-conjugated goat anti-mouse IgG second antibody was used for detection.

To generate recombinant protein antiserum, E. coli BL21(DE3)/pLysS was transformed with pHMW1-4, and expression of recombinant protein was induced with IPTG, as described above. A cell sonicate of the bacterial cells was prepared and separated into a supernatant and pellet fraction by centrifugation at 10,000 \times g for 30 min. The recombinant protein fractionated with the

pellet fraction. A rabbit was subcutaneously immunized on biweekly schedule with 1 mg of protein from the pellet fraction, the first dose given with Freund's complete adjuvant and subsequent doses with Freund's incomplete adjuvant. Following the fourth injection, the rabbit was bled. Prior to use in the Western blot assay, the antiserum was absorbed extensively with sonicates of the host E. coli strain transformed with cloning vector alone.

To assess the sharing of antigenic determinants between HMW1 and filamentous hemagglutinin, enzyme-linked immunosorbent assay (ELISA) plates (Costar, Cambridge, Mass.) were coated with 60 μ l of a 4-ug/ml solution of filamentous hemagglutinin in Dulbecco's phosphate-buffered saline per well for 2 h at room temperature. Wells were blocked for 1 h with 1% bovine serum albumin in Dulbecco's phosphate-buffered saline prior to addition of serum dilutions. rHMW1 antiserum was serially diluted in 0.1% Brij (Sigma, St. Louis, Mo.) in Dulbecco's phosphate-buffered saline and incubated for 3 h at room temperature. After being washed, the plates were incubated with peroxidase-conjugated goat anti-rabbit IgG antibody (Bio-Rad) for 2 h at room temperature and subsequently developed with 2,2'-azino-bis(3-ethylbenzthiazoline-6-sulfonic acid) (Sigma) at a concentration of 0.54 mg/ml in 0.1 M sodium citrate buffer, pH 4.2, containing 0.03% H_2O_2 . Absorbances were read on an automated ELISA reader.

Recombinant phage expressing HMW1 or HMW2 were recovered as follows. The non-typeable H. influenzae strain 12 genomic library was screened for clones expressing high-molecular-weight proteins with an E. coli-absorbed human serum sample containing a high titer of antibodies directed against the high-molecular-weight proteins.

Numerous strongly reactive clones were identified along with more weakly reactive ones. Twenty strongly reactive clones were plaque-purified and examined by Western blot for expression of recombinant proteins.

5 Each of the strongly reactive clones expressed one of two types of high-molecular-weight proteins, designated HMW1 and HMW2. The major immunoreactive protein bands in the HMW1 and HMW2 lysates migrated with apparent molecular masses of 125 and 120 kDa, respectively. In addition to the major bands, each lysate contained minor protein bands of higher apparent molecular weight. Protein bands seen in the HMW2 lysates at molecular masses of less than 120 kDa were not regularly observed and presumably represent proteolytic degradation products. Lysates of

10 LE392 infected with the λ EMBL3 cloning vector alone were non-reactive when immunologically screened with the same serum sample. Thus, the observed activity was not due to cross-reactive *E. coli* proteins or λ EMBL3-encoded proteins.

15 Furthermore, the recombinant proteins were not simply binding immunoglobulin nonspecifically, since the proteins were not reactive with the goat anti-human IgG conjugate alone, with normal rabbit sera, or with serum from a number of healthy young infants.

20 Representative clones expressing either the HMW1 or HMW2 recombinant proteins were characterized further. The restriction maps of the two phage types were different from each other, including the regions encoding the HMW1 and HMW2 structural genes. Figure 5A shows restriction maps of representative recombinant phage which contained the HMW1 or HMW2 structural genes. The locations of the structural genes are indicated by the shaded bars.

25 HMW1 plasmid subclones were constructed by using the T7 expression plasmid T7-7 (Fig. 5A and B). HMW2 plasmid subclones also were constructed, and the results with

these latter subclones were similar to those observed with the HMW1 constructs.

The approximate location and direction of transcription of the HMW1 structure gene were initially determined by using plasmid pHMW1 (Fig. 5A). This plasmid was constructed by inserting the 8.5-kb BamHI-SalI fragment from λ HMW1 into BamHI- and SalI-cut pT7-7. E. coli transformed with pHMW1 expressed an immunoreactive recombinant protein with an apparent molecular mass of 115 kDa, which was strongly inducible with IPTG. This protein was significantly smaller than the 125-kDa major protein expressed by the parent phage, indicating that it either was being expressed as a fusion protein or was truncated at the carboxy terminus.

To more precisely localize the 3' end of the structural gene, additional plasmids were constructed with progressive deletions from the 3' end of the pHMW1 construct. Plasmid pHMW1-1 was constructed by digestion of pHMW1 with PstI, isolation of the resulting 8.8-kb fragment, and religation. Plasmid pHMW1-2 was constructed by digestion of pHMW1 with HindIII, isolation of the resulting 7.5-kb fragment, and religation. E. coli transformed with either plasmid pHMW1-1 or pHMW1-2 also expressed an immunoreactive recombinant protein with an apparent molecular mass of 115 kDa. These results indicated that the 3' end of the structural gene was 5' of the HindIII site.

To more precisely localize the 5' end of the gene, plasmids pHMW1-4 and pHMW1-7 were constructed. Plasmid pHMW1-4 was constructed by cloning the 5.1-kb BamHI-HindIII fragment from λ HMW1 into a pT7-7-derived plasmid containing the upstream 3.8-kb EcoRI-BamHi fragment. E. coli transformed with pHMW1-4 expressed an immunoreactive protein with an apparent molecular mass of approximately 160 kDa. Although protein production was inducible with IPTG, the levels of protein production in these

transformants were substantially lower than those with the pHMW1-2 transformants described above. Plasmid pHMW1-7 was constructed by digesting pHMW1-4 with NdeI and SpeI. The 9.0-kbp fragment generated by this double digestion was isolated, blunt ended, and religated. *E. coli* transformed with pHMW1-7 also expressed an immunoreactive protein with an apparent molecular mass of 160 kDa, a protein identical in size to that expressed by the pHMW1-4 transformants. The result indicated that the initiation codon for the HMW1 structural gene was 3' of the SpeI site. DNA sequence analysis confirmed this conclusion.

As noted above, the λ HMW1 phage clones expressed a major immunoreactive band of 125 kDa, whereas the HMW1 plasmid clones pHMW1-4 and pHMW1-7, which contained what was believed to be the full-length gene, expressed an immunoreactive protein of approximately 160 kDa. This size discrepancy was disconcerting. One possible explanation was that an additional gene or genes necessary for correct processing of the HMW1 gene product were deleted in the process of subcloning. To address this possibility, plasmid pHMW1-14 was constructed. This construct was generated by digesting pHMW1 with NdeI and MluI and inserting the 7.6-kbp NdeI-MluI fragment isolated from pHMW1-4. Such a construct would contain the full-length HMW1 gene as well as the DNA 3' of the HMW1 gene which was present in the original HMW1 phage. *E. coli* transformed with this plasmid expressed major immunoreactive proteins with apparent molecular masses of 125 and 160 kDa as well as additional degradation products. The 125- and 160-kDa bands were identical to the major and minor immunoreactive bands detected in the HMW1 phage lysates. Interestingly, the pHMW1-14 construct also expressed significant amounts of protein in the uninduced condition, a situation not observed with the earlier constructs.

5 The relationship between the 125- and 160-kDa proteins remains somewhat unclear. Sequence analysis, described below, reveals that the HMW1 gene would be predicted to encode a protein of 159 kDa. It is believed that the 160-kDa protein is a precursor form of the mature 125-kDa protein, with the conversion from one protein to the other being dependent on the products of the two downstream genes.

10 Sequence analysis of the HMW1 gene (Figure 1) revealed a 4,608-bp open reading frame (ORF), beginning with an ATG codon at nucleotide 351 and ending with a TAG stop codon at nucleotide 4959. A putative ribosome-binding site with the sequence AGGAG begins 10 bp upstream of the putative initiation codon. Five other in-frame ATG codons are located within 250 bp of the beginning of the ORF, but none of these is preceded by a typical ribosome-binding site. The 5'-flanking region of the ORF contains a series of direct tandem repeats, with the 7-bp sequence ATCTTTC repeated 16 times. These tandem repeats stop 100 bp 5' of the putative initiation codon. An 8-bp inverted repeat characteristic of a rho-independent transcriptional terminator is present, beginning at nucleotide 4983, 25 bp 3' of the presumed translational stop. Multiple termination codons are present in all three reading frames both upstream and downstream of the ORF. The derived amino acid sequence of the protein encoded by the HMW1 gene (Figure 2) has a molecular weight of 159,000, in good agreement with the apparent molecular weights of the proteins expressed by the HMW1-4 and HMW1-7 transformants. The derived amino acid sequence of the amino terminus does not demonstrate the characteristics of a typical signal sequence. The BamHI site used in generation of pHMW1 comprises bp 1743 through 1748 of the nucleotide sequence. The ORF downstream of the BamHI site would be predicted to encode a protein of 111 kDa, in good agreement with the 115 kDa

estimated for the apparent molecular mass of the pHMW1-encoded fusion protein.

The sequence of the HMW2 gene (Figure 3) consists of a 4,431-bp ORF, beginning with an ATG codon at nucleotide 5 352 and ending with a TAG stop codon at nucleotide 4783. The first 1,259 bp of the ORF of the HMW2 gene are identical to those of the HMW1 gene. Thereafter, the sequences begin to diverge but are 80% identical overall. With the exception of a single base addition at 10 nucleotide 93 of the HMW2 sequence, the 5'-flanking regions of the HMW1 and HMW2 genes are identical for 310 bp upstream from the respective initiation codons. Thus, the HMW2 gene is preceded by the same set of tandem repeats and the same putative ribosome-binding site which 15 lies 5' of the HMW1 gene. A putative transcriptional terminator identical to that identified 3' of the HMW1 ORF is noted, beginning at nucleotide 4804. The discrepancy in the lengths of the two genes is principally accounted for by a 186-bp gap in the HMW2 sequence, beginning at nucleotide position 3839. The 20 derived amino acid sequence of the protein encoded by the HMW2 gene (Figure 4) has a molecular weight of 155,000 and is 71% identical with the derived amino acid sequence of the HMW1 gene.

25 The derived amino acid sequences of both the HMW1 and HMW2 genes (Figures 2 and 4) demonstrated sequence similarity with the derived amino acid sequence of filamentous hemagglutinin of Bordetella pertussis, a surface-associated protein of this organism. The initial 30 and optimized TFASTA scores for the HMW1-filamentous hemagglutinin sequence comparison were 87 and 186, respectively, with a word size of 2. The z score for the comparison was 45.8. The initial and optimized TFASTA scores for the HMW2-filamentous hemagglutinin sequence comparison were 68 and 196, respectively. The z score 35 for the latter comparison was 48.7. The magnitudes of

the initial and optimized TFASTA scores and the z scores suggested that a biologically significant relationship existed between the HMW1 and HMW2 gene products and filamentous hemagglutinin. When the derived amino acid sequences of HMW1, HMW2, and filamentous hemagglutinin genes were aligned and compared, the similarities were most notable at the amino-terminal ends of the three sequences. Twelve of the first 22 amino acids in the predicted peptide sequences were identical. In addition, the sequences demonstrated a common five-amino-acid stretch, Asn-Pro-Asn-Gly-Ile, and several shorter stretches of sequence identity within the first 200 amino acids.

Example 2:

To further explore the HMW1-filamentous hemagglutinin relationship, the ability of antiserum prepared against the HMW1-4 recombinant protein (rHMW1) to recognize purified filamentous hemagglutinin was assessed. The rHMW1 antiserum demonstrated ELISA reactivity with filamentous hemagglutinin in a dose-dependent manner. Preimmune rabbit serum had minimal reactivity in this assay. The rHMW1 antiserum also was examined in a Western blot assay and demonstrated weak but positive reactivity with purified filamentous hemagglutinin in this system also.

To identify the native Haemophilus protein corresponding to the HMW1 gene product and to determine the extent to which proteins antigenically related to the HMW1 cloned gene product were common among other non-typeable H. influenzae strains, a panel of Haemophilus strains was screened by Western blot with the rHMW1 antiserum. The antiserum recognized both a 125- and a 120-kDa protein band in the homologous strain 12, the putative mature protein products of the HMW1 and HMW2 genes, respectively.

When used to screen heterologous non-typeable H. influenzae strains, rHMW1 antiserum recognized high-molecular-weight proteins in 75% of 125 epidemiologically unrelated strains. In general, the antiserum reacted with one or two protein bands in the 100- to 150-kDa range in each of the heterologous strains in a pattern similar but not identical to that seen in the homologous strain.

Monoclonal antibody X3C is a murine IgG antibody directed against the filamentous hemagglutinin protein of B. pertussis. This antibody can inhibit the binding of B. pertussis cells to Chinese hamster ovary cells and HeLa cells in culture and will inhibit hemagglutination of erythrocytes by purified filamentous hemagglutinin. A Western blot assay was performed in which this monoclonal antibody was screened against the same panel of non-typeable H. influenzae strains discussed above. Monoclonal antibody X3C recognized both the high-molecular-weight proteins in non-typeable H. influenzae strain 12 which were recognized by the recombinant-protein antiserum. In addition, the monoclonal antibody recognized protein bands in a subset of heterologous non-typeable H. influenzae strains which were identical to those recognized by the recombinant-protein antiserum. On occasion, the filamentous hemagglutinin monoclonal antibody appeared to recognize only one of the two bands which had been recognized by the recombinant-protein antiserum. Overall, monoclonal antibody X3C recognized high-molecular-weight protein bands identical to those recognized by the rHMW1 antiserum in approximately 35% of our collection of non-typeable H. influenzae strains.

Example 3:

Mutants deficient in expression of HMW1, MW2 or both proteins were constructed to examine the role of these proteins in bacterial adherence. The following strategy was employed. pHMW1-14 (see Example 1, Figure 5A) was

digested with BamHI and then ligated to a kanamycin cassette isolated on a 1.3-kb BamHI fragment from pUC4K. The resultant plasmid (pHMW1-17) was linearized by digestion with XbaI and transformed into non-typeable H. influenzae strain 12, followed by selection for kanamycin resistant colonies. Southern analysis of a series of these colonies demonstrated two populations of transformants, one with an insertion in the HMW1 structural gene and the other with an insertion in the HMW2 structural gene. One mutant from each of these classes was selected for further studies.

Mutants deficient in expression of both proteins were recovered using the following protocol. After deletion of the 2.1-kb fragment of DNA between two EcoRI sites spanning the 3'-portion of the HMW1 structural gene in pHMW-15, the kanamycin cassette from pUC4K was inserted as a 1.3-kb EcoR1 fragment. The resulting plasmid (pHMW1-16) was linearized by digestion with XbaI and transformed into strain 12, followed again by selection for kanamycin resistant colonies. Southern analysis of a representative sampling of these colonies demonstrated that in seven of eight cases, insertion into both the HMW1 and HMW2 loci had occurred. One such mutant was selected for further studies.

To confirm the intended phenotypes, the mutant strains were examined by Western blot analysis with a polyclonal antiserum against recombinant HMW1 protein. The parental strain expressed both the 125-kD HMW1 and the 120-kD HMW2 protein. In contrast, the HMW2 mutant failed to express the 120-kD protein, and the HMW1 mutant failed to express the 125-kD protein. The double mutant lacked expression of either protein. On the basis of whole cell lysates, outer membrane profiles, and colony morphology, the wild type strain and the mutants were otherwise identical with one another. Transmission

electron microscopy demonstrated that none of the four strains expressed pili.

The capacity of wild type strain 12 to adhere to Chang epithelial cells was examined. In such assays, 5 bacteria were inoculated into broth and allowed to grow to a density of $\sim 2 \times 10^9$ cfu/ml. Approximately 2×10^7 cfu were inoculated onto epithelial cell monolayers, and plates were gently centrifuged at $165 \times g$ for 5 minutes to facilitate contact between bacteria and the epithelial 10 surface. After incubation for 30 minutes at $37^\circ C$ in 5% CO_2 , monolayers were rinsed 5 times with PBS to remove nonadherent organisms and were treated with trypsin-EDTA (0.05% trypsin, 0.5% EDTA) in PBS to release them from the plastic support. Well contents were agitated, and 15 dilutions were plated on solid medium to yield the number of adherent bacteria per monolayer. Percent adherence was calculated by dividing the number of adherent cfu per monolayer by the number of inoculated cfu.

As depicted in Table 1 below (the Tables appear at 20 the end of the descriptive text), this strain adhered quite efficiently, with nearly 90% of the inoculum binding to the monolayer. Adherence by the mutant expressing HMW1 but not HMW2 (HMW2⁻) was also quite efficient and comparable to that by the wild type strain. 25 In contrast, attachment by the strain expressing HMW2 but deficient in expression of HMW1 (HMW1⁻) was decreased about 15-fold relative to the wild type. Adherence by the double mutant (HMW1⁻/HMW2⁻) was decreased even further, approximately 50-fold compared with the wild 30 type and approximately 3-fold compared with the HMW1 mutant. Considered together, these results suggest that both the HMW1 protein and the, HMW2 protein influence attachment to Chang epithelial cells. Interestingly, 35 optimal adherence to this cell line appears to require HMW1 but not HMW2.

Example 4:

Using the plasmids pHMW1-16 and pHMW1-17 (see Example 3) and following a scheme similar to that employed with strain 12 as described in Example 3, three 5 non-typeable Haemophilus strain 5 mutants were isolated, including one with the kanamycin gene inserted into the hmw1-like (designated hmw3) locus, a second with an insertion in the hmw2-like (designated hmw4) locus, and a third with insertions in both loci. As predicted, 10 Western immunoblot analysis demonstrated that the mutant with insertion of the kanamycin cassette into the hmw1-like locus had lost expression of the HMW3 125-kD protein, while the mutant with insertion into the hmw2-like locus failed to express the HMW4 123-kD protein. 15 The mutant with a double insertion was unable to express either of the high molecular weight proteins.

As shown in Table 1 below, wild type strain 5 demonstrated high level adherence, with almost 80% of the inoculum adhering per monolayer. Adherence by the mutant 20 deficient in expression of the HMW2-like protein was also quite high. In contrast, adherence by the mutant unable to express the HMW1-like protein was reduced about 5-fold relative to the wild type, and attachment by the double mutant was diminished even further (approximately 25-fold). Examination of Giemsa-stained samples 25 confirmed these observations (not shown). Thus, the results with strain 5 corroborate the findings with strain 12 and the HMW1 and HMW2 proteins.

Example 5:

To confirm an adherence function for the HMW1 and HMW2 proteins and to examine the effect of HMW1 and HMW2 independently of other H. influenzae surface structures, the hmw1 and the hmw2 gene clusters were introduced into E. coli DH5 α , using plasmids pHMW1-14 and pHMW2-21, respectively. As a control, the cloning vector, pT7-7, 35 was also transformed into E. coli DH5 α . Western blot

analysis demonstrated that E. coli DH5 α containing the hmw1 genes expressed a 125 kDa protein, while the same strain harboring the hmw2 genes expressed a 120-kDa protein. E. coli DH5 α containing pT7-7 failed to react with antiserum against recombinant HMW1. Transmission electron microscopy revealed no pili or other surface appendages on any of the E. coli strains.

Adherence by the E. coli strains was quantitated and compared with adherence by wild type non-typeable H. influenzae strain 12. As shown in Table 2 below, adherence by E. coli DH5 α containing vector alone was less than 1% of that for strain 12. In contrast, E. coli DH5 α harboring the hmw1 gene cluster demonstrated adherence levels comparable to those for strain 12. Adherence by E. coli DH5 α containing the hmw2 genes was approximately 6-fold lower than attachment by strain 12 but was increased 20-fold over adherence by E. coli DH5 α with pT7-7 alone. These results indicate that the HMW1 and HMW2 proteins are capable of independently mediating attachment to Chang conjunctival cells. These results are consistent with the results with the H. influenzae mutants reported in Examples 3 and 4, providing further evidence that, with Chang epithelial cells, HMW1 is a more efficient adhesin than is HMW2.

Experiments with E. coli HB101 harboring pT7-7, pHMW1-14, or pHMW2-21 confirmed the results obtained with the DH5 α derivatives (see Table 2).

Example 6:

HMW1 and HMW2 were isolated and purified from non-typeable H. influenzae (NTHI) strain 12 in the following manner. Non-typeable Haemophilus bacteria from frozen stock culture were streaked onto a chocolate plate and grown overnight at 37°C in an incubator with 5% CO₂. 50ml starter culture of brain heart infusion (BHI) broth, supplemented with 10 μ g/ml each of hemin and NAD was inoculated with growth on chocolate plate. The starter

5 culture was grown until the optical density (O.D. - 600nm) reached 0.6 to 0.8 and then the bacteria in the starter culture was used to inoculate six 500 ml flasks of supplemented BHI using 8 to 10 ml per flask. The bacteria were grown in 500 ml flasks for an additional 5 to 6 hours at which time the O.D. was 1.5 or greater. Cultures were centrifuged at 10,000 rpm for 10 minutes.

10 Bacterial pellets were resuspended in a total volume of 250 ml of an extraction solution comprising 0.5 M NaCl, 0.01 M Na₂EDTA, 0.01 M Tris 50 μ M 1,10-phenanthroline, pH 7.5. The cells were not sonicated or otherwise disrupted. The resuspended cells were allowed to sit on ice at 0°C for 60 minutes. The resuspended cells were centrifuged at 10,000 rpm for 10 minutes at 15 4°C to remove the majority of intact cells and cellular debris. The supernatant was collected and centrifuged at 100,000 xg for 60 minutes at 4°C. The supernatant again was collected and dialyzed overnight at 4°C against 0.01 M sodium phosphate, pH 6.0.

20 The sample was centrifuged at 10,000 rpm for 10 minutes at 4°C to remove insoluble debris precipitated from solution during dialysis. The supernatant was applied to a 10 ml CM Sepharose column which has been pre-equilibrated with 0.01 M sodium phosphate, pH 6. 25 Following application to this column, the column was washed with 0.01 M sodium phosphate. Proteins were elevated from the column with a 0 - 0.5M KCl gradient in 0.01 M Na phosphate, pH 6 and fractions were collected for gel examination. Coomassie gels of column fractions 30 were carried out to identify those fractions containing high molecular weight proteins. The fractions containing high molecular weight proteins were pooled and concentrated to a 1 to 3 ml volume in preparation for application of sample to gel filtration column.

35 A Sepharose CL-4B gel filtration column was equilibrated with phosphate-buffered saline, pH 7.5. The

concentrated high molecular weight protein sample was applied to the gel filtration column and column fractions were collected. Coomassie gels were performed on the column fractions to identify those containing high molecular weight proteins. The column fractions containing high molecular weight proteins were pooled.

5 The proteins were tested to determine whether they would protect against experimental otitis media caused by the homologous strain.

10 Chinchillas received three monthly subcutaneous injections with 40 μ g of an HMW1-HMW2 protein mixture in Freund's adjuvant. One month after the last injection, the animals were challenged by intrabullar inoculation with 300 cfu of NTHI strain 12.

15 Infection developed in 5 of 5 control animals versus 5 of 10 immunized animals. Among infected animals, geometric mean bacterial counts in middle ear fluid 7 days post-challenge were 7.4×10^6 in control animals versus 1.3×10^5 in immunized animals.

20 Serum antibody titres following immunization were comparable in uninfected and infected animals. However, infection in immunized animals was uniformly associated with the appearance of bacteria down-regulated in expression of the HMW proteins, suggesting bacterial 25 selection in response to immunologic pressure.

30 Although this data shows that protection following immunization was not complete, this data suggests the HMW adhesin proteins are potentially important protective antigens which may comprise one component of a multi-component NTHI vaccine.

35 These animal challenge tests were repeated in Chinchillas at a lower dose challenge than the 300 cfu employed above. In this instance, complete protection was achieved. In these experiments, groups of five animals were immunized with 20 μ g of the HMW1-HMW2

5 mixture on days 1, 28, and 42 in the presence of AlPO₄. Blood samples were collected on day 53 to monitor the antibody response. On day 56, the left ear of animals was challenged with about 10 cfu of H. influenzae strain 12. Ear infection was monitored on day 4. Four animals in Group 3 were infected previously by H. influenzae strain 12 and were recovered completely for at least one month before the second challenge. The results are outlined in the following Table A:

10

TABLE A

15

Protective ability of HMW protein against non-typeable H. influenzae challenge in chinchilla model

20

Group (#)	Antigens	Total Animals	Number of Animals Showed Positive Ear Infection		
			Tympano- gram	Otosco- pic Examina- tion	cfu of Bac- teria/ 10 μ L
1	HMW	5	0	0	0
2	None	5	5	5	850- 3200 (4/5)
3	Convalescent	4	0	0	0

25

Example 7:

30

A number of synthetic peptides were derived from HMW1. Antisera then was raised to these peptides. The anti-peptide antisera to peptide HMW1-P5 was shown to recognize HMW1. Peptide HMW1-P5 covers amino acids 1453 to 1481 of HMW1, has the sequence VDEVIEAKRILEKVKDLSDEEREALAKLG (SEQ ID NO:9), and represents bases 1498 to 1576 in Figure 10.

35

This finding demonstrates that the DNA sequence and the derived protein is being interpreted in the correct

reading frame and that peptides derived from the sequence can be produced which will be immunogenic.

SUMMARY OF DISCLOSURE

5 In summary of this disclosure, the present invention provides high molecular weight proteins of non-typeable Haemophilus, genes coding for the same and vaccines incorporating such proteins. Modifications are possible within the scope of this invention.

Table 1. Effect of mutation of high molecular weight proteins on adherence to Chang epithelial cells by nontypable *H. influenzae*.

<u>Strain</u>	<u>ADHERENCE*</u>	
	<u>\odot inoculum</u>	<u>relative to wild type†</u>
Strain 12 derivatives		
wild type	87.7 \pm 5.9	100.0 \pm 6.7
HMW1- mutant	6.0 \pm 0.9	6.8 \pm 1.0
HMW2- mutant	89.9 \pm 10.8	102.5 \pm 12.3
HMW1-/HMW2- mutant	2.0 \pm 0.3	2.3 \pm 0.3
Strain 5 derivatives		
wild type	78.7 \pm 3.2	100.0 \pm 4.1
HMW1-like mutant	15.7 \pm 2.6	19.9 \pm 3.3
HMW2-like mutant	103.7 \pm 14.0	131.7 \pm 17.8
double mutant	3.5 \pm 0.6	4.4 \pm 0.8

* Numbers represent mean (\pm standard error of the mean) of measurements in triplicate or quadruplicate from representative experiments.

† Adherence values for strain 12 derivatives are relative to strain 12 wild type; values for strain 5 derivatives are relative to strain 5 wild type.

- 27 -

Table 2. Adherence by *E. coli* DH5 α and HB101 harboring *hmw1* or *hmw2* gene clusters.

<u>Strain*</u>	Adherence relative to <i>H. influenzae</i> strain 12 [†]
DH5 α (pT7-7)	0.7 \pm 0.02
DH5 α (pHMW1-14)	114.2 \pm 15.9
DH5 α (pHMW2-21)	14.0 \pm 3.7
HB101 (pT7-7)	1.2 \pm 0.5
HB101 (pHMW1-14)	93.6 \pm 15.8
HB101 (pHMW2-21)	3.6 \pm 0.9

* The plasmid pHMW1-14 contains the *hmw1* gene cluster, while pHMW2-21 contains the *hmw2* gene cluster; pT7-7 is the cloning vector used in these constructs.

† Numbers represent the mean (\pm standard error of the mean) of measurements made in triplicate from representative experiments.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: BARENKAMP, STEPHEN J
ST. GEME III, JOSEPH W
- (ii) TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
OF NON-TYPEABLE HAEMOPHILUS

(iii) NUMBER OF SEQUENCES: 8

(iv) CORRESPONDENCE ADDRESS:

- (A) ADDRESSEE: Shoemaker and Mattare, Ltd
- (B) STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
Bldg. 1
- (C) CITY: Arlington
- (D) STATE: Virginia
- (E) COUNTRY: U.S.A.
- (F) ZIP: 22202-0286

(v) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) OPERATING SYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:

- (A) APPLICATION NUMBER: US 08/038,682
- (B) FILING DATE: 16-MAR-1993
- (C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:

- (A) NAME: BERKSTRESSER, JERRY W
- (B) REGISTRATION NUMBER: 22,651
- (C) REFERENCE/DOCKET NUMBER: 1038-293

(ix) TELECOMMUNICATION INFORMATION:

- (A) TELEPHONE: (703) 415-0810
- (B) TELEFAX: (703) 415-0813

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 5116 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ACAGCGTTCT CTTAATACTA GTACAAACCC ACAATAAAAT ATGACAAACA ACAATTACAA	60
CACCTTTTTT GCAGTCTATA TGCAAATATT TTAAAAAAATA GTATAAATCC GCCATATAAA	120
ATGGTATAAT CTTTCATCTT TCATCTTCA TCTTCATCT TTCATCTTTC ATCTTCATC	180
TITCATCTTT CATCTTCAT CTTTCATCTT TCATCTTCA TCTTCATCT TTCATCTTTC	240
ACATGCCCTG ATGAACCGAG GGAAGGGAGG GAGGGGCAAG AATGAAGAGG GAGCTGAACG	300

SUBSTITUTE SHEET (RULE 26)

AACGCAAATG ATAAAGTAAT TTAATTGTTCAACTAACCTT AGGAGAAAAT ATGAACAAAGC	360
TATATCGTCT CAAATTCAAGC AAACGCCTGA ATGCTTGGT TGCTGTGTCT GAATTGGCAC	420
GGGGTTGTGA CCATTCCACA GAAAAAGGCA GCGAAAACC TGCTCGCATG AAAGTGCCTC	480
ACTTAGCGTT AAAGCCACTT TCCGCTATGT TACTATCTTT AGGTGTAACA TCTATTCCAC	540
AATCTGTTT AGCAAGCGGC TTACAAGGAA TGGATGTAGT ACACGGCACA GCCACTATGC	600
AAGTAGATGG TAATAAAACC ATTATCCGCA ACAGTGTGA CGATATCATT AATTGGAAAC	660
AATTTAACAT CGACCAAAAT GAAATGGTGC AGTTTTTACA AGAAAACAAC AACTCCGCCG	720
TATTCAACCG TGTTACATCT AACCAAATCT CCCAATTAAA AGGGATTTA GATTCTAACG	780
GACAAGTCTT TTTAATCAAC CCAAATGGTA TCACAATAGG TAAAGACGCA ATTATTAACA	840
CTAATGGCTT TACGGCTTCT ACGCTAGACA TTTCTAACGA AAACATCAAG GCGCGTAATT	900
TCACCTTCGA GCAAACAAA GATAAAGCGC TCGCTGAAAT TGTGAATCAC GGTTTAATT	960
CTGTCGGTAA AGACGGCAGT GTAAATCTTA TTGGTGGCAA AGTAAAAAAC GAGGGTGTGA	1020
TTAGCGTAAA TGGTGGCAGC ATTCTTTAC TCGCAGGGCA AAAATCACC ATCAGCGATA	1080
TAATAAACCC AACCATTACT TACAGCATTG CCGCGCCTGA AAATGAAGCG GTCAATCTGG	1140
GCGATATTT TGCCAAAGGC GGTAACATTA ATGTCCGTGC TGCCACTATT CGAAACCAAG	1200
GTAAACTTTC TGCTGATTCT GTAAGCAAAG ATAAAAGCGG CAATATTGTT CTTTCCGCCA	1260
AAGAGGGTGA AGCGGAAATT GGCGGTGTAA TTTCCGCTCA AAATCAGCAA GCTAAAGGCG	1320
GCAAGCTGAT GATTACAGGC GATAAAGTCA CATTAAAAC AGGTGCAGTT ATCGACCTTT	1380
CAGGTAAGA AGGGGGAGAA ACTTACCTTG GCGGTGACGA GCGCGGCGAA GGTAAAAGG	1440
GCATTCAATT AGCAAAGAAA ACCTCTTAG AAAAAGGCTC AACCATCAAT GTATCAGGCA	1500
AAGAAAAGG CGGACGCGCT ATTGTGTGGG GCGATATTGC GTTAATTGAC GGCAATATTA	1560
ACGCTCAAGG TAGTGGTGAT ATCGCTAAA CCGGTGGTT TGTGGAGACG TCAGGGCATG	1620
ATTTATTCAAT CAAAGACAAT GCAATTGTTG ACGCCAAAGA GTGGTTGTTA GACCCGGATA	1680
ATGTATCTAT TAATGCAGAA ACAGCAGGAC GCAGCAATAC TTCAGAAGAC GATGAATACA	1740
CGGGATCCGG GAATAGTGCC AGCACCCCAA AACGAAACAA AGAAAAGACA ACATTAACAA	1800
ACACAACCTCT TGAGAGTATA CTAAAAAAAG GTACCTTGT TAACATCACT GCTAATCAAC	1860
GCATCTATGT CAATAGCTCC ATTAATTTAT CCAATGGCAG CTTAACTCTT TGGAGTGAGG	1920
GTCGGAGCGG TGGCGCGTT GAGATTAACA ACGATATTAC CACCGGTGAT GATACCAGAG	1980
GTGCAAACCTT ACAATTTAC TCAGGCGGCT GGGTTGATGT TCATAAAAAT ATCTCACTCG	2040
GGGCGCAAGG TAACATAAAC ATTACAGCTA AACAAAGATAT CGCCTTGAG AAAGGAAGCA	2100
ACCAAGTCAT TACAGGTCAA GGGACTATTAA CCTCAGGCAA TCAAAAAGGT TTTAGATTAA	2160
ATAATGTCTC TCTAAACGGC ACTGGCAGCG GACTGCAATT CACCACTAAA AGAACCAATA	2220
AATACGCTAT CACAAATAAA TTTGAAGGGA CTTTAAATAT TTCAGGGAAA GTGAACATCT	2280
CAATGGTTT ACCTAAAAAT GAAAGTGGAT ATGATAAATT CAAAGGACGC ACTTACTGGA	2340

ATTTAACCTC	CTTAAATGTT	TCCGAGAGTG	GCGAGTTAA	CCTCACTATT	GAETCCAGAG	2400
GAAGCGATAG	TGCAGGCACA	CTTACCCAGC	CTTATAATTT	AAACGGTATA	TCATTCAACA	2460
AAGACACTAC	CTTTAATGTT	GAACGAAATG	CAAGAGTCAA	CTTGACATC	AAGGCACCAA	2520
TAGGGATAAA	TAAGTATTCT	AGTTTGAATT	ACGCATCATT	TAATGGAAAC	ATTCAGTT	2580
CGGGAGGGGG	GAGTGTGAT	TTCACACTTC	TCGCCTCATC	CTCTAACGTC	CAAACCCCCG	2640
GTGTAGTTAT	AAATTCTAAA	TACTTTAATG	TTTCAACAGG	GTCAAGTTA	AGATTAAAAA	2700
CTTCAGGCTC	AACAAAAACT	GGCTTCTCAA	TAGAGAAAGA	TTTAACTTA	AATGCCACCG	2760
GAGGCAACAT	AACACTTTG	CAAGTTGAAG	GCACCCGATGG	AATGATTGGT	AAAGGCATTG	2820
TAGCCAAAAAA	AAACATAACC	TTTGAAGGAG	GTAACATCAC	CTTGGCTCC	AGGAAAGCCG	2880
TAACAGAAAT	CGAAGGCAAT	GTTACTATCA	ATAACAACGC	TAACGTCACT	CTTATCGGTT	2940
CGGATTTGA	CAACCATCAA	AAACCTTTAA	CTATTAAAAA	AGATGTCATC	ATTAATAGCG	3000
GCAACCTTAC	CGCTGGAGGC	AATATTGTCA	ATATAGCCGG	AAATCTTACC	GTTGAAAGTA	3060
ACGCTAATTT	CAAAGCTATC	ACAAATTCA	CTTTAATGT	AGGCGGCTTG	TTTGACAACA	3120
AAGGCAATT	AAATATTCC	ATTGCCAAAG	GAGGGGCTCG	CTTAAAGAC	ATTGATAATT	3180
CCAAGAATT	AAGCATCACC	ACCAACTCCA	GCTCCACTTA	CCGCACTATT	ATAAGCGGCA	3240
ATATAACCAA	TAAAAACGGT	GATTTAAATA	TTACGAACGA	AGGTAGTGAT	ACTGAAATGC	3300
AAATTGGCGG	CGATGTCTCG	CAAAAAGAAG	GTAATCTCAC	GATTTCTTCT	GACAAAATCA	3360
ATATTACCAA	ACAGATAACA	ATCAAGGCAG	GTGTTGATGG	GGAGAATTCC	GATTCAAGACG	3420
CGACAAACAA	TGCCAATCTA	ACCATTAAAA	CCAAAGAATT	GAAATTAACG	CAAGACCTAA	3480
ATATTCAGG	TTCAATAAA	GCAGAGATTA	CAGCTAAAGA	TGGTAGTGAT	TTAACTATTG	3540
GTAACACCAA	TAGTGCTGAT	GGTACTAATG	CCAAAAAAAGT	AACTTTAAC	CAGGTTAAAG	3600
ATTCAAAAT	CTCTGCTGAC	GGTCACAAGG	TGACACTACA	CAGCAAAGTG	GAAACATCCG	3660
GTAGTAATAA	CAACACTGAA	GATAGCAGTG	ACAATAATGC	CGGCTTAACT	ATCGATGCAA	3720
AAAATGTAAC	AGTAAACAAC	AATATTACTT	CTCACAAAGC	AGTGAGCATC	TCTGCGACAA	3780
GTGGAGAAAT	TACCACTAAA	ACAGGTACAA	CCATTAACGC	AACCACGTG	AACGTGGAGA	3840
TAACCGCTCA	AACAGGTAGT	ATCCTAGGTG	GAATTGAGTC	CAGCTCTGGC	TCTGTAACAC	3900
TTACTGCAAC	CGAGGGCGCT	CTTGCTGTA	GCAATATTTC	GGGCAACACC	GTTACTGTTA	3960
CTGCAAATAG	CGGTGCATTA	ACCACTTGG	CAGGCTCTAC	AATTAAAGGA	ACCGAGAGTG	4020
TAACCACTTC	AAGTCAATCA	GGCGATATCG	GCGGTACGAT	TTCTGGTGGC	ACAGTAGAGG	4080
TTAAAGCAAC	CGAAAGTTA	ACCACTCAAT	CCAATTCAA	AATTAAAGCA	ACAACAGGCG	4140
AGGCTAACGT	ACAAAGTGCA	ACAGGTACAA	TTGGTGGTAC	GATTTCCGGT	AATACGGTAA	4200
ATGTTACGGC	AAACGCTGGC	GATTTAACAG	TTGGGAATGG	CGCAGAAATT	AATGCGACAG	4260
AAGGAGCTGC	AACCTTAAC	ACATCATCGG	GCAAATTAAAC	TACCGAAGCT	AGTTCACACA	4320
TTACTTCAGC	CAAGGGTCAG	GTAAATCTTT	CAGCTCAGGA	TGGTAGCGTT	GCAGGAAGTA	4380

CONSTITUTE SHEET (RULE 26)

TTAATGCCGC CAATGTGACA CTAATACTA CAGGCACTTT AACTACCGTG AAGGGTTCAA	4440
ACATTAATGC AACCAGCGGT ACCTTGGTTA TTAACGCAA AGACGCTGAG CTAAATGGCG	4500
CAGCATTGGG TAACCACACA GTGGTAAATG CAACCAACGC AAATGGCTCC GGCAGCGTAA	4560
TCGCGACAAC CTCAAGCAGA GTGAACATCA CTGGGGATTT AATCACAATA AATGGATTAA	4620
ATATCATTTC AAAAAACGGT ATAAACACCG TACTGTTAAA AGGCGTTAAA ATTGATGTGA	4680
AATACATTCA ACCGGGTATA GCAAGCGTAG ATGAAGTAAT TGAAGCGAAA CGCATTCTTG	4740
AGAAGGTAAA AGATTTATCT GATGAAGAAA GAGAAGCGTT AGCTAAACTT GGAGTAAGTG	4800
CTGTACGTTT TATTGAGCCA AATAATACAA TTACAGTCGA TACACAAAAT GAATTTGCAA	4860
CCAGACCATT AAGTCGAATA GTGATTTCTG AAGGCAGGGC GTGTTCTCA AACAGTGATG	4920
GCGCGACGGT GTGCGTTAAT ATCGCTGATA ACGGGCGGTA GCGGTCAGTA ATTGACAAGG	4980
TAGATTCAT CCTGCAATGA AGTCATTTA TTTTCGTATT ATTTACTGTG TGGGTTAAAG	5040
TTCAGTACGG GCTTTACCCA TCTTGTAAAA AATTACGGAG AATACAATAA AGTATTTTA	5100
ACAGGTTATT ATTATG	5116

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1536 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Asn Lys Ile Tyr Arg Leu Lys Phe Ser Lys Arg Leu Asn Ala Leu	
1 5 10 15	
Val Ala Val Ser Glu Leu Ala Arg Gly Cys Asp His Ser Thr Glu Lys	
20 25 30	
Gly Ser Glu Lys Pro Ala Arg Met Lys Val Arg His Leu Ala Leu Lys	
35 40 45	
Pro Leu Ser Ala Met Leu Leu Ser Leu Gly Val Thr Ser Ile Pro Gln	
50 55 60	
Ser Val Leu Ala Ser Gly Leu Gln Gly Met Asp Val Val His Gly Thr	
65 70 75 80	
Ala Thr Met Gln Val Asp Gly Asn Lys Thr Ile Ile Arg Asn Ser Val	
85 90 95	
Asp Ala Ile Ile Asn Trp Lys Gln Phe Asn Ile Asp Gln Asn Glu Met	
100 105 110	
Val Gln Phe Leu Gln Glu Asn Asn Ser Ala Val Phe Asn Arg Val	
115 120 125	
Thr Ser Asn Gln Ile Ser Gln Leu Lys Gly Ile Leu Asp Ser Asn Gly	
130 135 140	

Gln Val Phe Leu Ile Asn Pro Asn Gly Ile Thr Ile Gly Lys Asp Ala
 145 150 155 160
 Ile Ile Asn Thr Asn Gly Phe Thr Ala Ser Thr Leu Asp Ile Ser Asn
 165 170 175
 Glu Asn Ile Lys Ala Arg Asn Phe Thr Phe Glu Gln Thr Lys Asp Lys
 180 185 190
 Ala Leu Ala Glu Ile Val Asn His Gly Leu Ile Thr Val Gly Lys Asp
 195 200 205
 Gly Ser Val Asn Leu Ile Gly Gly Lys Val Lys Asn Glu Gly Val Ile
 210 215 220
 Ser Val Asn Gly Gly Ser Ile Ser Leu Leu Ala Gly Gln Lys Ile Thr
 225 230 235 240
 Ile Ser Asp Ile Ile Asn Pro Thr Ile Thr Tyr Ser Ile Ala Ala Pro
 245 250 255
 Glu Asn Glu Ala Val Asn Leu Gly Asp Ile Phe Ala Lys Gly Gly Asn
 260 265 270
 Ile Asn Val Arg Ala Ala Thr Ile Arg Asn Gln Gly Lys Leu Ser Ala
 275 280 285
 Asp Ser Val Ser Lys Asp Lys Ser Gly Asn Ile Val Leu Ser Ala Lys
 290 295 300
 Glu Gly Glu Ala Glu Ile Gly Gly Val Ile Ser Ala Gln Asn Gln Gln
 305 310 315 320
 Ala Lys Gly Gly Lys Leu Met Ile Thr Gly Asp Lys Val Thr Leu Lys
 325 330 335
 Thr Gly Ala Val Ile Asp Leu Ser Gly Lys Glu Gly Gly Glu Thr Tyr
 340 345 350
 Leu Gly Gly Asp Glu Arg Gly Glu Gly Lys Asn Gly Ile Gln Leu Ala
 355 360 365
 Lys Lys Thr Ser Leu Glu Lys Gly Ser Thr Ile Asn Val Ser Gly Lys
 370 375 380
 Glu Lys Gly Gly Arg Ala Ile Val Trp Gly Asp Ile Ala Leu Ile Asp
 385 390 395 400
 Gly Asn Ile Asn Ala Gln Gly Ser Gly Asp Ile Ala Lys Thr Gly Gly
 405 410 415
 Phe Val Glu Thr Ser Gly His Asp Leu Phe Ile Lys Asp Asn Ala Ile
 420 425 430
 Val Asp Ala Lys Glu Trp Leu Leu Asp Phe Asp Asn Val Ser Ile Asn
 435 440 445
 Ala Glu Thr Ala Gly Arg Ser Asn Thr Ser Glu Asp Asp Glu Tyr Thr
 450 455 460
 Gly Ser Gly Asn Ser Ala Ser Thr Pro Lys Arg Asn Lys Glu Lys Thr
 465 470 475 480
 Thr Leu Thr Asn Thr Thr Leu Glu Ser Ile Leu Lys Lys Gly Thr Phe
 485 490 495

Val Asn Ile Thr Ala Asn Gln Arg Ile Tyr Val Asn Ser Ser Ile Asn
 500 505 510
 Leu Ser Asn Gly Ser Leu Thr Leu Trp Ser Glu Gly Arg Ser Gly Gly
 515 520 525
 Gly Val Glu Ile Asn Asn Asp Ile Thr Thr Gly Asp Asp Thr Arg Gly
 530 535 540
 Ala Asn Leu Thr Ile Tyr Ser Gly Gly Trp Val Asp Val His Lys Asn
 545 550 555 560
 Ile Ser Leu Gly Ala Gln Gly Asn Ile Asn Ile Thr Ala Lys Gln Asp
 565 570 575
 Ile Ala Phe Glu Lys Gly Ser Asn Gln Val Ile Thr Gly Gln Gly Thr
 580 585 590
 Ile Thr Ser Gly Asn Gln Lys Gly Phe Arg Phe Asn Asn Val Ser Leu
 595 600 605
 Asn Gly Thr Gly Ser Gly Leu Gln Phe Thr Thr Lys Arg Thr Asn Lys
 610 615 620
 Tyr Ala Ile Thr Asn Lys Phe Glu Gly Thr Leu Asn Ile Ser Gly Lys
 625 630 635 640
 Val Asn Ile Ser Met Val Leu Pro Lys Asn Glu Ser Gly Tyr Asp Lys
 645 650 655
 Phe Lys Gly Arg Thr Tyr Trp Asn Leu Thr Ser Leu Asn Val Ser Glu
 660 665 670
 Ser Gly Glu Phe Asn Leu Thr Ile Asp Ser Arg Gly Ser Asp Ser Ala
 675 680 685
 Gly Thr Leu Thr Gln Pro Tyr Asn Leu Asn Gly Ile Ser Phe Asn Lys
 690 695 700
 Asp Thr Thr Phe Asn Val Glu Arg Asn Ala Arg Val Asn Phe Asp Ile
 705 710 715 720
 Lys Ala Pro Ile Gly Ile Asn Lys Tyr Ser Ser Leu Asn Tyr Ala Ser
 725 730 735
 Phe Asn Gly Asn Ile Ser Val Ser Gly Gly Ser Val Asp Phe Thr
 740 745 750
 Leu Leu Ala Ser Ser Ser Asn Val Gln Thr Pro Gly Val Val Ile Asn
 755 760 765
 Ser Lys Tyr Phe Asn Val Ser Thr Gly Ser Ser Leu Arg Phe Lys Thr
 770 775 780
 Ser Gly Ser Thr Lys Thr Gly Phe Ser Ile Glu Lys Asp Leu Thr Leu
 785 790 795 800
 Asn Ala Thr Gly Gly Asn Ile Thr Leu Leu Gln Val Glu Gly Thr Asp
 805 810 815
 Gly Met Ile Gly Lys Gly Ile Val Ala Lys Lys Asn Ile Thr Phe Glu
 820 825 830
 Gly Gly Asn Ile Thr Phe Gly Ser Arg Lys Ala Val Thr Glu Ile Glu
 835 840 845

SUBSTITUTE SHEET (RULE 26)

Gly Asn Val Thr Ile Asn Asn Ala Asn Val Thr Leu Ile Gly Ser
 850 855 860
 Asp Phe Asp Asn His Gln Lys Pro Leu Thr Ile Lys Lys Asp Val Ile
 865 870 875 880
 Ile Asn Ser Gly Asn Leu Thr Ala Gly Gly Asn Ile Val Asn Ile Ala
 885 890 895
 Gly Asn Leu Thr Val Glu Ser Asn Ala Asn Phe Lys Ala Ile Thr Asn
 900 905 910
 Phe Thr Phe Asn Val Gly Gly Leu Phe Asp Asn Lys Gly Asn Ser Asn
 915 920 925
 Ile Ser Ile Ala Lys Gly Gly Ala Arg Phe Lys Asp Ile Asp Asn Ser
 930 935 940
 Lys Asn Leu Ser Ile Thr Thr Asn Ser Ser Thr Tyr Arg Thr Ile
 945 950 955 960
 Ile Ser Gly Asn Ile Thr Asn Lys Asn Gly Asp Leu Asn Ile Thr Asn
 965 970 975
 Glu Gly Ser Asp Thr Glu Met Gln Ile Gly Gly Asp Val Ser Gln Lys
 980 985 990
 Glu Gly Asn Leu Thr Ile Ser Ser Asp Lys Ile Asn Ile Thr Lys Gln
 995 1000 1005
 Ile Thr Ile Lys Ala Gly Val Asp Gly Glu Asn Ser Asp Ser Asp Ala
 1010 1015 1020
 Thr Asn Asn Ala Asn Leu Thr Ile Lys Thr Lys Glu Leu Lys Leu Thr
 1025 1030 1035 1040
 Gln Asp Leu Asn Ile Ser Gly Phe Asn Lys Ala Glu Ile Thr Ala Lys
 1045 1050 1055
 Asp Gly Ser Asp Leu Thr Ile Gly Asn Thr Asn Ser Ala Asp Gly Thr
 1060 1065 1070
 Asn Ala Lys Lys Val Thr Phe Asn Gln Val Lys Asp Ser Lys Ile Ser
 1075 1080 1085
 Ala Asp Gly His Lys Val Thr Leu His Ser Lys Val Glu Thr Ser Gly
 1090 1095 1100
 Ser Asn Asn Asn Thr Glu Asp Ser Ser Asp Asn Asn Ala Gly Leu Thr
 1105 1110 1115 1120
 Ile Asp Ala Lys Asn Val Thr Val Asn Asn Ile Thr Ser His Lys
 1125 1130 1135
 Ala Val Ser Ile Ser Ala Thr Ser Gly Glu Ile Thr Thr Lys Thr Gly
 1140 1145 1150
 Thr Thr Ile Asn Ala Thr Thr Gly Asn Val Glu Ile Thr Ala Gln Thr
 1155 1160 1165
 Gly Ser Ile Leu Gly Gly Ile Glu Ser Ser Ser Gly Ser Val Thr Leu
 1170 1175 1180
 Thr Ala Thr Glu Gly Ala Leu Ala Val Ser Asn Ile Ser Gly Asn Thr
 1185 1190 1195 1200

SUBSTITUTE SHEET (RULE 26)

Val Thr Val Thr Ala Asn Ser Gly Ala Leu Thr Thr Leu Ala Gly Ser
 1205 1210 1215
 Thr Ile Lys Gly Thr Glu Ser Val Thr Thr Ser Ser Gln Ser Gly Asp
 1220 1225 1230
 Ile Gly Gly Thr Ile Ser Gly Gly Thr Val Glu Val Lys Ala Thr Glu
 1235 1240 1245
 Ser Leu Thr Thr Gln Ser Asn Ser Lys Ile Lys Ala Thr Thr Gly Glu
 1250 1255 1260
 Ala Asn Val Thr Ser Ala Thr Gly Thr Ile Gly Gly Thr Ile Ser Gly
 1265 1270 1275 1280
 Asn Thr Val Asn Val Thr Ala Asn Ala Gly Asp Leu Thr Val Gly Asn
 1285 1290 1295
 Gly Ala Glu Ile Asn Ala Thr Glu Gly Ala Ala Thr Leu Thr Thr Ser
 1300 1305 1310
 Ser Gly Lys Leu Thr Thr Glu Ala Ser Ser His Ile Thr Ser Ala Lys
 1315 1320 1325
 Gly Gln Val Asn Leu Ser Ala Gln Asp Gly Ser Val Ala Gly Ser Ile
 1330 1335 1340
 Asn Ala Ala Asn Val Thr Leu Asn Thr Thr Gly Thr Leu Thr Thr Val
 1345 1350 1355 1360
 Lys Gly Ser Asn Ile Asn Ala Thr Ser Gly Thr Leu Val Ile Asn Ala
 1365 1370 1375
 Lys Asp Ala Glu Leu Asn Gly Ala Ala Leu Gly Asn His Thr Val Val
 1380 1385 1390
 Asn Ala Thr Asn Ala Asn Gly Ser Gly Ser Val Ile Ala Thr Thr Ser
 1395 1400 1405
 Ser Arg Val Asn Ile Thr Gly Asp Leu Ile Thr Ile Asn Gly Leu Asn
 1410 1415 1420
 Ile Ile Ser Lys Asn Gly Ile Asn Thr Val Leu Leu Lys Gly Val Lys
 1425 1430 1435 1440
 Ile Asp Val Lys Tyr Ile Gln Pro Gly Ile Ala Ser Val Asp Glu Val
 1445 1450 1455
 Ile Glu Ala Lys Arg Ile Leu Glu Lys Val Lys Asp Leu Ser Asp Glu
 1460 1465 1470
 Glu Arg Glu Ala Leu Ala Lys Leu Gly Val Ser Ala Val Arg Phe Ile
 1475 1480 1485
 Glu Pro Asn Asn Thr Ile Thr Val Asp Thr Gln Asn Glu Phe Ala Thr
 1490 1495 1500
 Arg Pro Leu Ser Arg Ile Val Ile Ser Glu Gly Arg Ala Cys Phe Ser
 1505 1510 1515 1520
 Asn Ser Asp Gly Ala Thr Val Cys Val Asn Ile Ala Asp Asn Gly Arg
 1525 1530 1535

SUBSTITUTE SHEET (RULE 26)

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4937 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

TAAATATACA	AGATAATAAA	AATAAATCAA	GATTTTGTG	ATGACAAACA	ACAATTACAA	60
CACCTTTTTT	GCAGTCTATA	TGCAAATATT	TTAAAAAAAT	AGTATAAATC	CGCCATATAA	120
AATGGTATAA	TCTTCATCT	TTCATCTTTA	ATCTTCATC	TTTCATCTTT	CATCTTCAT	180
CTTTCATCTT	TCATCTTC	TCTTCATCT	TTCATCTTC	ATCTTCATC	TTTCATCTTT	240
CACATGAAAT	GATGAACCGA	GGGAAGGGAG	GGAGGGCAA	GAATGAAGAG	GGAGCTGAAC	300
GAACGCAAAT	GATAAAAGTAA	TTTAATTGTT	CAACTAACCT	TAGGAGAAAA	TATGAACAAG	360
ATATATCGTC	TCAAATTCA	CAAACGCCTG	AATGCTTGG	TTGCTGTGTC	TGAATTGGCA	420
CGGGGTTGTG	ACCATTCCAC	AGAAAAAGGC	TTCCGCTATG	TTACTATCTT	TAGGTGTAAC	480
CACTTAGCGT	TAAAGCCACT	TTCCGCTATG	TTACTATCTT	TAGGTGTAAC	ATCTATTCCA	540
CAATCTGTT	TAGCAAGCGG	CTTACAAGGA	ATGGATGTAG	TACACGGCAC	AGCCACTATG	600
CAAGTAGATG	GTAATAAAAC	CATTATCCGC	AACAGTGTG	ACGCTATCAT	TAATTGGAAA	660
CAATTAAACA	TCGACCAAAA	TGAAATGGTG	CAGTTTTAC	AAGAAAACAA	CAACTCCGCC	720
GTATTCAACC	GTGTTACATC	TAACCAAATC	TCCCAATTAA	AAGGGATTTT	AGATTCTAAC	780
GGACAAGTCT	TTTTAATCAA	CCCAAATGGT	ATCACAATAG	GTAAAGACGC	AATTATTAAC	840
ACTAATGGCT	TTACGGCTTC	TACGCTAGAC	ATTCTAACG	AAAACATCAA	GGCGCGTAAT	900
TTCACCTTCG	AGCAAACCAA	AGATAAAAGCG	CTCGCTGAAA	TTGTGAATCA	CGGTTTAATT	960
ACTGTCGGTA	AAGACGGCAG	TGTAAATCTT	ATTGGTGGCA	AAGTAAAAAA	CGAGGGTGTG	1020
ATTAGCGTAA	ATGGTGGCAG	CATTTCTTTA	CTCGCAGGGC	AAAAAATCAC	CATCAGCGAT	1080
ATAATAAAACC	CAACCATTAC	TTACAGCATT	GCCGCGCCTG	AAAATGAAGC	GGTCAATCTG	1140
GGCGATATTT	TTGCCAAAGG	CGGTAACATT	AATGTCCGTG	CTGCCACTAT	TCGAAACCAA	1200
GGTAAACTTT	CTGCTGATT	TGTAAGCAA	GATAAAAGCG	GCAATATTGT	TCTTCCGCC	1260
AAAGAGGGTG	AAGCGGAAAT	TGGCGGTGTA	ATTTCCGCTC	AAAATCAGCA	AGCTAAAGGC	1320
GGCAAGCTGA	TGATTACAGG	CGATAAAAGTC	ACATTAACAAA	CAGGTGCAGT	TATCGACCTT	1380
TCAGGTAAAG	AAGGGGGAGA	AACTTACCTT	GGCGGTGACG	AGCGCGCGA	AGGTAAAAAC	1440
GGCATTCAAT	TAGCAAAGAA	AACCTCTTTA	AAAAAAGGCT	CAACCATCAA	TGTATCAGGC	1500
AAAGAAAAAG	GCGGACGCGC	TATTGTGTGG	GGCGATATTG	CGTTAATTGA	CGGCAATATT	1560
AACGCTCAAG	GTAGTGGTGA	TATCGCTAAA	ACCGGTGGTT	TTGTGGAGAC	ATCGGGGCAT	1620

SUBSTITUTE SHEET (RULE 26)

TATTTATCCA TTGACAGCAA TGCAATTGTT AAAACAAAAG AGTGGTTGCT AGACCCCTGAT	1680
GATGTAACAA TTGAAGCCGA AGACCCCCTT CGCAATAATA CCGGTATAAA TGATGAATT	1740
CCAACAGGCA CCGGTGAAGC AAGCGACCCCT AAAAAAAAATA GCGAACTCAA AACAAACGCTA	1800
ACCAATACAA CTATTTCAAA TTATCTGAAA AACGCCTGGA CAATGAATAT AACGGCATCA	1860
AGAAAACCTTA CCGTTAATAG CTCAATCAAC ATCGGAAGCA ACTCCCACCTT AATTCTCCAT	1920
AGTAAAGGTC AGCGTGGCGG AGGCCTTCAG ATTGATGGAG ATATTACTTC TAAAGGCGGA	1980
AATTTAACCA TTTATTCTGG CGGATGGGTT GATGTCATA AAAATATTAC GCTTGATCAG	2040
GGTTTTTAA ATATTACCGC CGCTTCCGTA GCTTTGAAG GTGGAAATAA CAAAGCACGC	2100
GACGCGGCAA ATGCTAAAAT TGTCGCCCAG GGCACTGTAA CCATTACAGG AGAGGGAAAA	2160
GATTCAGGG CTAACAACGT ATCTTTAAC GGAACGGGTAA AAGGTCTGAA TATCATTCA	2220
TCAGTGAATA ATTTAACCCA CAATCTTAGT GGCACAATTAA ACATATCTGG GAATATAACA	2280
ATTAACCAAA CTACGAGAAA GAACACCTCG TATTGGCAAA CCAGCCATGA TTCGCACTGG	2340
AACGTCAGTG CTCTTAATCT AGAGACAGGC GCAAATTAA CCTTTATTAA ATACATTCA	2400
AGCAATAGCA AAGGCTTAAC AACACAGTAT AGAAGCTCTG CAGGGGTGAA TTTAACGGC	2460
GTAAATGGCA ACATGTCATT CAATCTAAA GAAGGAGCGA AAGTTAATTAA CAAATTAAAA	2520
CCAAACGAGA ACATGAACAC AAGCAAACCT TTACCAATTG GGTGTTAGC CAATATCACA	2580
GCCACTGGTG GGGGCTCTGT TTTTTTGAT ATATATGCCA ACCATTCTGG CAGAGGGCT	2640
GAGTTAAAAA TGAGTGAAT TAATATCTCT AACGGCGCTA ATTTTACCTT AAATTCCCAT	2700
GTTCGGGCG ATGACGCTTT TAAAATCAAC AAAGACTTAA CCATAAATGC AACCAATTCA	2760
AATTCAGCC TCAGACAGAC GAAAGATGAT TTTTATGACG GGTACGCACG CAATGCCATC	2820
AATTCAACCT ACAACATATC CATTCTGGC GGTAATGTCA CCCTTGGTGG ACAAAACCTCA	2880
AGCAGCAGCA TTACGGGGAA TATTACTATC GAGAAAGCAG CAAATGTTAC GCTAGAAGCC	2940
AATAACGCC CTAATCAGCA AAACATAAGG GATAGAGTTA TAAAACCTGG CAGCTTGCTC	3000
GTTAATGGGA GTTTAAGTTT AACTGGCGAA AATGCAGATA TTAAAGGCAA TCTCACTATT	3060
TCAGAAAGCG CCACTTAA AGGAAAGACT AGAGATAACCC TAAATATCAC CGGCAATTAA	3120
ACCAATAATG GCACTGCCGA AATTAATATA ACACAAGGAG TGGTAAAAGT TGGCAATGTT	3180
ACCAATGATG GTGATTAAA CATTACCACT CACGCTAAAC GCAACCAAAG AAGCATCATC	3240
GGCGGAGATA TAATCAACAA AAAAGGAAGC TTAAATATTA CAGACAGTAA TAATGATGCT	3300
GAAATCCAAA TTGGCGGCAA TATCTCGCAA AAAGAAGGCA ACCTCACGAT TTCTTCCGAT	3360
AAAATTAAATA TCACCAAACA GATAACAATC AAAAAGGGTA TTGATGGAGA GGACTCTAGT	3420
TCAGATGCGA CAAGTAATGC CAACCTAACT ATTAAAACCA AAGAATTGAA ATTGACAGAA	3480
GACCTAAGTA TTTCAGGTTT CAATAAAGCA GAGATTACAG CCAAAGATGG TAGAGATTAA	3540
ACTATTGGCA ACAGTAATGA CGGTAACAGC GGTGCCGAAG CCAAAACAGT AACTTTAAC	3600
AATGTTAAAG ATTCAAAAT CTCTGCTGAC GGTACAATG TGACACTAAA TAGCAAAGTG	3660

AAAACATCTA	GCAGCAATGG	CGGACGTGAA	AGCAATAGCG	ACAACGATAAC	CGGCTTAAC	3720
ATTACTGCAA	AAAATGTAGA	AGTAAACAAA	GATATTACTT	CTCTAAAC	AGTAAATATC	3780
ACCGCGTCGG	AAAAGGTTAC	CACCACAGCA	GGCTCGACCA	TTAACGCAAC	AAATGGCAAA	3840
GCAAGTATTA	CAACCAAAAC	AGGTGATATC	AGCGGTACGA	TTTCCGGTAA	CACGGTAAGT	3900
GTTAGCGCGA	CTGGTGATTT	AACCCTAAA	TCCGGCTCAA	AAATTGAAGC	GAAATCGGGT	3960
GAGGCTAATG	TAACAAGTGC	AACAGGTACA	ATTGGCGGT	CAATTTCGG	TAATACGGTA	4020
AATGTTACGG	CAAACGCTGG	CGATTTAAC	GTTGGGAATG	GCGCAGAAAT	TAATGCGACA	4080
GAAGGAGCTG	CAACCTTAAC	CGCAACAGGG	AATACCTTGA	CTACTGAAGC	CGGTTCTAGC	4140
ATCACTTCAA	CTAAGGGTCA	GGTAGACCTC	TTGGCTCAGA	ATGGTAGCAT	CGCAGGAAGC	4200
ATTAATGCTG	CTAATGTGAC	ATTAATTAAC	ACAGGCACCT	TAACCACCGT	GGCAGGGCTCG	4260
GATATTAAAG	CAACCAGCGG	CACCTTGGTT	ATTAACGCAA	AAGATGCTAA	GCTAAATGGT	4320
GATGCATCAG	GTGATAGTAC	AGAAGTGAAT	GCAGTCAACG	CAAGCGGCTC	TGGTAGTGTG	4380
ACTCGGGCAA	CCTCAAGCAG	TGTGAATATC	ACTGGGGATT	TAAACACAGT	AAATGGGTTA	4440
AATATCATTT	CGAAAGATGG	TAGAAACACT	GTGCGCTTAA	GAGGCAAGGA	AATTGAGGTG	4500
AAATATATCC	AGCCAGGTGT	AGCAAGTGT	GAAGAAGTAA	TTGAAGCGAA	ACGCGTCCTT	4560
GAAAAAGTAA	AAGATTATAC	TGATGAAGAA	AGAGAAACAT	TAGCTAAACT	TGGTAGTAAGT	4620
GCTGTACGTT	TTGTTGAGCC	AAATAATACA	ATTACAGTCA	ATACACAAAA	TGAATTAC	4680
ACCAGACCGT	CAAGTCAAGT	GATAATTCT	GAAGGTAAGG	CGTGTTCCTC	AAGTGGTAAT	4740
GGCGCACGAG	TATGTACCAA	TGTTGCTGAC	GATGGACAGC	CGTAGTCAGT	AATTGACAAG	4800
GTAGATTCA	TCCTGCAATG	AAGTCATTTT	ATTTTCGTAT	TATTTACTGT	GTGGGTTAAA	4860
GTTCACTACG	GGCTTTACCC	ATCTTGTAAA	AAATTACGGA	GAATACAATA	AAGTATTTT	4920
AACAGGTTAT	TATTATG					4937

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1477 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met	Asn	Lys	Ile	Tyr	Arg	Leu	Lys	Phe	Ser	Lys	Arg	Leu	Asn	Ala	Leu
1				5					10						15
Val	Ala	Val	Ser	Glu	Leu	Ala	Arg	Gly	Cys	Asp	His	Ser	Thr	Glu	Lys
				20				25							30
Gly	Ser	Glu	Lys	Pro	Ala	Arg	Met	Lys	Val	Arg	His	Leu	Ala	Leu	Lys
				35				40							45

Pro Leu Ser Ala Met Leu Leu Ser Leu Gly Val Thr Ser Ile Pro Gln
 50 55 60

Ser Val Leu Ala Ser Gly Leu Gln Gly Met Asp Val Val His Gly Thr
 65 70 75 80

Ala Thr Met Gln Val Asp Gly Asn Lys Thr Ile Ile Arg Asn Ser Val
 85 90 95

Asp Ala Ile Ile Asn Trp Lys Gln Phe Asn Ile Asp Gln Asn Glu Met
 100 105 110

Val Gln Phe Leu Gln Glu Asn Asn Ser Ala Val Phe Asn Arg Val
 115 120 125

Thr Ser Asn Gln Ile Ser Gln Leu Lys Gly Ile Leu Asp Ser Asn Gly
 130 135 140

Gln Val Phe Leu Ile Asn Pro Asn Gly Ile Thr Ile Gly Lys Asp Ala
 145 150 155 160

Ile Ile Asn Thr Asn Gly Phe Thr Ala Ser Thr Leu Asp Ile Ser Asn
 165 170 175

Glu Asn Ile Lys Ala Arg Asn Phe Thr Phe Glu Gln Thr Lys Asp Lys
 180 185 190

Ala Leu Ala Glu Ile Val Asn His Gly Leu Ile Thr Val Gly Lys Asp
 195 200 205

Gly Ser Val Asn Leu Ile Gly Gly Lys Val Lys Asn Glu Gly Val Ile
 210 215 220

Ser Val Asn Gly Gly Ser Ile Ser Leu Leu Ala Gly Gln Lys Ile Thr
 225 230 235 240

Ile Ser Asp Ile Ile Asn Pro Thr Ile Thr Tyr Ser Ile Ala Ala Pro
 245 250 255

Glu Asn Glu Ala Val Asn Leu Gly Asp Ile Phe Ala Lys Gly Asn
 260 265 270

Ile Asn Val Arg Ala Ala Thr Ile Arg Asn Gln Gly Lys Leu Ser Ala
 275 280 285

Asp Ser Val Ser Lys Asp Lys Ser Gly Asn Ile Val Leu Ser Ala Lys
 290 295 300

Glu Gly Glu Ala Glu Ile Gly Gly Val Ile Ser Ala Gln Asn Gln Gln
 305 310 315 320

Ala Lys Gly Gly Lys Leu Met Ile Thr Gly Asp Lys Val Thr Leu Lys
 325 330 335

Thr Gly Ala Val Ile Asp Leu Ser Gly Lys Glu Gly Gly Glu Thr Tyr
 340 345 350

Leu Gly Gly Asp Glu Arg Gly Glu Gly Lys Asn Gly Ile Gln Leu Ala
 355 360 365

Lys Lys Thr Ser Leu Glu Lys Gly Ser Thr Ile Asn Val Ser Gly Lys
 370 375 380

Glu Lys Gly Gly Phe Ala Ile Val Trp Gly Asp Ile Ala Leu Ile Asp
 385 390 395 400

40

Gly Asn Ile Asn Ala Gln Gly Ser Gly Asp Ile Ala Lys Thr Gly Gly
 405 410 415
 Phe Val Glu Thr Ser Gly His Asp Leu Phe Ile Lys Asp Asn Ala Ile
 420 425 430
 Val Asp Ala Lys Glu Trp Leu Leu Asp Phe Asp Asn Val Ser Ile Asn
 435 440 445
 Ala Glu Asp Pro Leu Phe Asn Asn Thr Gly Ile Asn Asp Glu Phe Pro
 450 455 460
 Thr Gly Thr Gly Glu Ala Ser Asp Pro Lys Lys Asn Ser Glu Leu Lys
 465 470 475 480
 Thr Thr Leu Thr Asn Thr Thr Ile Ser Asn Tyr Leu Lys Asn Ala Trp
 485 490 495
 Thr Met Asn Ile Thr Ala Ser Arg Lys Leu Thr Val Asn Ser Ser Ile
 500 505 510
 Asn Ile Gly Ser Asn Ser His Leu Ile Leu His Ser Lys Gly Gln Arg
 515 520 525
 Gly Gly Gly Val Gln Ile Asp Gly Asp Ile Thr Ser Lys Gly Gly Asn
 530 535 540
 Leu Thr Ile Tyr Ser Gly Gly Trp Val Asp Val His Lys Asn Ile Thr
 545 550 555 560
 Leu Asp Gln Gly Phe Leu Asn Ile Thr Ala Ala Ser Val Ala Phe Glu
 565 570 575
 Gly Gly Asn Asn Lys Ala Arg Asp Ala Ala Asn Ala Lys Ile Val Ala
 580 585 590
 Gln Gly Thr Val Thr Ile Thr Gly Glu Gly Lys Asp Phe Arg Ala Asn
 595 600 605
 Asn Val Ser Leu Asn Gly Thr Gly Lys Gly Leu Asn Ile Ile Ser Ser
 610 615 620
 Val Asn Asn Leu Thr His Asn Leu Ser Gly Thr Ile Asn Ile Ser Gly
 625 630 635 640
 Asn Ile Thr Ile Asn Gln Thr Thr Arg Lys Asn Thr Ser Tyr Trp Gln
 645 650 655
 Thr Ser His Asp Ser His Trp Asn Val Ser Ala Leu Asn Leu Glu Thr
 660 665 670
 Gly Ala Asn Phe Thr Phe Ile Lys Tyr Ile Ser Ser Asn Ser Lys Gly
 675 680 685
 Leu Thr Thr Gln Tyr Arg Ser Ser Ala Gly Val Asn Phe Asn Gly Val
 690 695 700
 Asn Gly Asn Met Ser Phe Asn Leu Lys Glu Ala Lys Val Asn Phe
 705 710 715 720
 Lys Leu Lys Pro Asn Glu Asn Met Asn Thr Ser Lys Pro Leu Pro Ile
 725 730 735
 Arg Phe Leu Ala Asn Ile Thr Ala Thr Gly Gly Ser Val Phe Phe
 740 745 750

SUBSTITUTE SHEET (RULE 26)

Asp Ile Tyr Ala Asn His Ser Gly Arg Gly Ala Glu Leu Lys Met Ser
 755 760 765
 Glu Ile Asn Ile Ser Asn Gly Ala Asn Phe Thr Leu Asn Ser His Val
 770 775 780
 Arg Gly Asp Asp Ala Phe Lys Ile Asn Lys Asp Leu Thr Ile Asn Ala
 785 790 795 800
 Thr Asn Ser Asn Phe Ser Leu Arg Gln Thr Lys Asp Asp Phe Tyr Asp
 805 810 815
 Gly Tyr Ala Arg Asn Ala Ile Asn Ser Thr Tyr Asn Ile Ser Ile Leu
 820 825 830
 Gly Gly Asn Val Thr Leu Gly Gly Gln Asn Ser Ser Ser Ile Thr
 835 840 845
 Gly Asn Ile Thr Ile Glu Lys Ala Ala Asn Val Thr Leu Glu Ala Asn
 850 855 860
 Asn Ala Pro Asn Gln Gln Asn Ile Arg Asp Arg Val Ile Lys Leu Gly
 865 870 875 880
 Ser Leu Leu Val Asn Gly Ser Leu Ser Leu Thr Gly Glu Asn Ala Asp
 885 890 895
 Ile Lys Gly Asn Leu Thr Ile Ser Glu Ser Ala Thr Phe Lys Gly Lys
 900 905 910
 Thr Arg Asp Thr Leu Asn Ile Thr Gly Asn Phe Thr Asn Asn Gly Thr
 915 920 925
 Ala Glu Ile Asn Ile Thr Gln Gly Val Val Lys Leu Gly Asn Val Thr
 930 935 940
 Asn Asp Gly Asp Leu Asn Ile Thr Thr His Ala Lys Arg Asn Gln Arg
 945 950 955 960
 Ser Ile Ile Gly Gly Asp Ile Ile Asn Lys Lys Gly Ser Leu Asn Ile
 965 970 975
 Thr Asp Ser Asn Asn Asp Ala Glu Ile Gln Ile Gly Gly Asn Ile Ser
 980 985 990
 Gln Lys Glu Gly Asn Leu Thr Ile Ser Ser Asp Lys Ile Asn Ile Thr
 995 1000 1005
 Lys Gln Ile Thr Ile Lys Lys Gly Ile Asp Gly Glu Asp Ser Ser Ser
 1010 1015 1020
 Asp Ala Thr Ser Asn Ala Asn Leu Thr Ile Lys Thr Lys Glu Leu Lys
 1025 1030 1035 1040
 Leu Thr Glu Asp Leu Ser Ile Ser Gly Phe Asn Lys Ala Glu Ile Thr
 1045 1050 1055
 Ala Lys Asp Gly Arg Asp Leu Thr Ile Gly Asn Ser Asn Asp Gly Asn
 1060 1065 1070
 Ser Gly Ala Glu Ala Lys Thr Val Thr Phe Asn Asn Val Lys Asp Ser
 1075 1080 1085
 Lys Ile Ser Ala Asp Gly His Asn Val Thr Leu Asn Ser Lys Val Lys
 1090 1095 1100

SUBSTITUTE SHEET (RULE 26)

Thr Ser Ser Ser Asn Gly Gly Arg Glu Ser Asn Ser Asp Asn Asp Thr
 1105 1110 1115 1120
 Gly Leu Thr Ile Thr Ala Lys Asn Val Glu Val Asn Lys Asp Ile Thr
 1125 1130 1135
 Ser Leu Lys Thr Val Asn Ile Thr Ala Ser Glu Lys Val Thr Thr Thr
 1140 1145 1150
 Ala Gly Ser Thr Ile Asn Ala Thr Asn Gly Lys Ala Ser Ile Thr Thr
 1155 1160 1165
 Lys Thr Gly Asp Ile Ser Gly Thr Ile Ser Gly Asn Thr Val Ser Val
 1170 1175 1180
 Ser Ala Thr Val Asp Leu Thr Thr Lys Ser Gly Ser Lys Ile Glu Ala
 1185 1190 1195 1200
 Lys Ser Gly Glu Ala Asn Val Thr Ser Ala Thr Gly Thr Ile Gly Gly
 1205 1210 1215
 Thr Ile Ser Gly Asn Thr Val Asn Val Thr Ala Asn Ala Gly Asp Leu
 1220 1225 1230
 Thr Val Gly Asn Gly Ala Glu Ile Asn Ala Thr Glu Gly Ala Ala Thr
 1235 1240 1245
 Leu Thr Ala Thr Gly Asn Thr Leu Thr Thr Glu Ala Gly Ser Ser Ile
 1250 1255 1260
 Thr Ser Thr Lys Gly Gln Val Asp Leu Leu Ala Gln Asn Gly Ser Ile
 1265 1270 1275 1280
 Ala Gly Ser Ile Asn Ala Ala Asn Val Thr Leu Asn Thr Thr Gly Thr
 1285 1290 1295
 Leu Thr Thr Val Ala Gly Ser Asp Ile Lys Ala Thr Ser Gly Thr Leu
 1300 1305 1310
 Val Ile Asn Ala Lys Asp Ala Lys Leu Asn Gly Asp Ala Ser Gly Asp
 1315 1320 1325
 Ser Thr Glu Val Asn Ala Val Asn Ala Ser Gly Ser Gly Ser Val Thr
 1330 1335 1340
 Ala Ala Thr Ser Ser Val Asn Ile Thr Gly Asp Leu Asn Thr Val
 1345 1350 1355 1360
 Asn Gly Leu Asn Ile Ile Ser Lys Asp Gly Arg Asn Thr Val Arg Leu
 1365 1370 1375
 Arg Gly Lys Glu Ile Glu Val Lys Tyr Ile Gln Pro Gly Val Ala Ser
 1380 1385 1390
 Val Glu Glu Val Ile Glu Ala Lys Arg Val Leu Glu Lys Val Lys Asp
 1395 1400 1405
 Leu Ser Asp Glu Glu Arg Glu Thr Leu Ala Lys Leu Gly Val Ser Ala
 1410 1415 1420
 Val Arg Phe Val Glu Pro Asn Asn Thr Ile Thr Val Asn Thr Gln Asn
 1425 1430 1435 1440
 Glu Phe Thr Thr Arg Pro Ser Ser Gln Val Ile Ile Ser Glu Gly Lys
 1445 1450 1455

SUBSTITUTE SHEET (RULE 26)

Ala Cys Phe Ser Ser Gly Asn Gly Ala Arg Val Cys Thr Asn Val Ala
 1460 1465 1470

Asp Asp Gly Gln Pro
 1475

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9171 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

ACAGCGTTCT CTTAATACCA	GTACAAACCC ACAATAAAAT ATGACAAACA ACAATTACAA	60
CACCTTTTTT GCAGTCTATA TGCAAATATT TTAAAAAATA GTATAAATCC GCCATATAAA		120
ATGGTATAAT CTTTCATCTT TCATCTTC CA TCTTCATCTT TTCA TCTTCATC ATCTTCATC		180
TTTCATCTT CATCTTCAT CTTTCATCTT TCATCTTC CA TCTTCATCTT TTCA TCTTCATC		240
ACATGAAATG ATGAACCGAG GGAAGGGAGG GAGGGGCAAG AATGAAGAGG GAGCTGAACG		300
AACGCAAATG ATAAAGTAAT TTAATTGTTT AACTAACCTT AGGAGAAAAT ATGAACAAGA		360
TATATCGTCT CAAATTCA GC AAACGCCTGA ATGCTTTGGT TGCTGTGTCT GAATTGGCAC		420
GGGGTTGTGA CCATTCCACA GAAAAAGGCA GCGAAAAACC TGCTCGCATG AAAGTGCCTC		480
ACTTAGCGTT AAAGCCACTT TCCGCTATGT TACTATCTT AGGTGTAACA TCTATTCCAC		540
AATCTGTTT AGCAAGCGGC TTACAAGGAA TGGATGTAGT ACACGGCACA GCCACTATGC		600
AAGTAGATGG TAATAAAACC ATTATCCGCA ACAGTGTGCA CGCTATCATT AATTGGAAAC		660
AATTTAACAT CGACCAAAAT GAAATGGTGC AGTTTTTACA AGAAAACAAC AACTCCGCCG		720
TATTCAACCG TGTACATCT AACCAAATCT CCCAATTAAA AGGGATTTA GATTCTAACG		780
GACAAGTCTT TTTAATCAAC CCAAATGGTA TCACAATAGG TAAAGACGCA ATTATTAACA		840
CTAATGGCTT TACGGCTTCT ACGCTAGACA TTTCTAACGA AAACATCAAG GCGCGTAATT		900
TCACCTTCGA GCAAACCAAA GATAAAGCGC TCGCTGAAAT TGTGAATCAC GGTTAATTA		960
CTGTCGGTAA AGACGGCAGT GTAAATCTT TTGGTGGCAA AGTAAAAAC GAGGGTGTGA		1020
TTAGCGTAA TGGTGGCAGC ATTTCTTAC TCGCAGGGCA AAAATCACC ATCAGCGATA		1080
TAATAAAACCC AACCATTACT TACAGCATTG CCGCGCCTGA AAATGAAGCG GTCAATCTGG		1140
GCGATATTTT TGCCAAAGGC GGTAACATTA ATGTCCGTGC TGCCACTATT CGAAACCAAG		1200
CTTTCCGCCA AAGAGGGTGA AGCGGAAATT GGCGGTGTAA TTTCCGCTCA AAATCAGCAA		1260
GCTAAAGGCG GCAAGCTGAT GATTACAGGC GATAAAAGTCA CATTAAAAAC AGGTGCAGTT		1320
ATCGACCTTT CAGGTAAAGA AGGGGGAGAA ACTTACCTTG GCGGTGACGA GCGCGCGAA		1380
GGTAAAAACG GCATTCAATT AGCAAAGAAA ACCTCTTGT AGGAGGCTC AACCATCAAT		1440

SUBSTITUTE SHEET (RULE 26)

GTATCAGGCA AAGAAAAAGG CGGACGCGCT ATTGTGTGGG GCGATATTGC GTTAATTGAC	1500
GGCAATATTA ACGCTCAAGG TAGTGGTGAT ATCGCTAAAA CCGGTGGTTT TGTGGAGACG	1560
TCGGGGCATG ATTTATTCCAT CAAAGACAAT GCAATTGTTG ACGCCAAAGA GTGGTTGTTA	1620
GACCCGGATA ATGTATCTAT TAATCCAGAA ACAGCAGGAC GCAGCAATAC TTCAGAAGAC	1680
GATGAATACA CGGGATCCGG GAATAGTGCC AGCACCCCAA AACGAAAACAA AGAAAAGACA	1740
ACATTAACAA ACACAACCTCT TGAGAGTATA CTAAAAAAAG GTACCTTGTTA TAACATCACT	1800
GCTAATCAAC GCATCTATGT CAATAGCTCC ATTAATTAT CCAATGGCAG CTTAACTCTT	1860
TGGAGTGAGG GTCGGAGCGG TGGCGGCCTT GAGATTAACA ACGATATTAC CACCGGTGAT	1920
GATACCAGAG GTGCAAACCTT ACAAAATTAC TCAGGCGGCT GGGTTGATGT TCATAAAAAT	1980
ATCTCACTCG GGGCGCAAGG TAACATAAAC ATTACAGCTA ACAAGATAT CGCCTTGAG	2040
AAAGGAAGCA ACCAAGTCAT TACAGGTCAA GGGACTATTAA CCTCAGGCAA TCAAAAAGGT	2100
TTTAGATTAA ATAATGTCTC TCTAAACGGC ACTGGCAGCG GACTGCAATT CACCACTAAA	2160
AGAACCAATA AATACGCTAT CACAAATAAA TTTGAAGGGA CTTTAAATAT TTCAGGGAAA	2220
GTGAACATCT CAATGGTTTT ACCTAAAAAT GAAAGTGGAT ATGATAAAATT CAAAGGACGC	2280
ACTTACTGGA ATTTAACCTC GAAAGTGGAT ATGATAAAATT CAAAGGACGC CCTCACTATT	2340
GAECTCAGAG GAAGCGATAG TGCAGGCACA CTTACCCAGC CTTATAATTAAACCGGTATA	2400
TCATTCAACA AAGACACTAC CTTAATGTT GAACGAAATG CAAGAGTCAG CTTTGACATC	2460
AAGGCACCAA TAGGGATAAA TAAGTATTCT AGTTTGAATT ACGCATCATT TAATGGAAAC	2520
ATTCAGTTT CGGGAGGGGG GAGTGGTGTAT TTCACACTTC TCGCCTCATC CTCTAACGTC	2580
CAAACCCCG GTGTAGTTAT AAATTCTAAA TACTTTAATG TTTCAACAGG GTCAAGTTA	2640
AGATTTAAA CTTCAGGCTC AACAAAAACT GGCTCTCAA TAGAGAAAGA TTTAACTTTA	2700
AATGCCACCG GAGGCAACAT AACACTTTG CAAGTTGAAG GCACCGATGG AATGATTGGT	2760
AAAGGCATTG TAGCCAAAAA AAACATAACC TTTGAAGGAG GTAAGATGAG GTTTGGCTCC	2820
AGGAAAGCCG TAACAGAAAT CGAAGGCAAT GTTACTATCA ATAACAAACGC TAACGTCACT	2880
CTTATCGGTT CGGATTTGA CAACCCTCAA AAACCTTAA CTATTAAAAA AGATGTCATC	2940
ATTAATAGCG GCAACCTTAC CGCTGGAGGC AATATTGTCA ATATAGCCGG AAATCTTACC	3000
GTTGAAAGTA ACGCTAATT CAAAGCTATC ACAAAATTCA CTTTTAATGT AGGCGGTTG	3060
TTTGACAACA AAGGCAATTCA AAATATTTC ATTGCCAAAG GAGGGGCTCG CTTTAAAGAC	3120
ATTGATAATT CCAAGAATT AAGCATCACC ACCAACTCCA GCTCCACTTA CGCAGTATT	3180
ATAAGCGGCA ATATAACCAA TAAAAACGGT GATTAAATA TTACGAACGA AGGTAGTGAT	3240
ACTGAAATGC AAATTGGCGG CGATGTCTCG CAAAAAGAAG GTAATCTCAC GATTCTTCT	3300
GACAAAATCA ATATTACCAA ACAGATAACA ATCAAGGCAG GTGTTGATGG GGAGAATTCC	3360
GATTCAAGACG CGACAAACAA TGCCAATCTA ACCATTAAAA CCAAAGAATT GAAATTAAACG	3420
CAAGACCTAA ATATTCAGG TTTCAATAAA GCAGAGATTA CAGCTAAAGA TGGTAGTGAT	3480

TTAACTATTG	GTAACACCAA	TAGTGCTGAT	GGTACTAATG	CCAAAAAAGT	AACCTTTAAC	3540
CAGGTTAAAG	ATTCAAAAAT	CTCTGCTGAC	GGTCACAAGG	TGACACTACA	CAGCAAAGTG	3600
GAACATCCG	GTAGTAATAA	CAACACTGAA	GATAGCAGTG	ACAATAATGC	CGGCTTAACT	3660
ATCGATGCAA	AAAATGTAAC	AGTAAACAAC	AATATTACTT	CTCACAAAGC	AGTGAGCATC	3720
TCTGCGACAA	GTGGAGAAAT	TACCACTAAA	ACAGGTACAA	CCATTAACGC	AACCACGTGGT	3780
AACGTGGAGA	TAACCGCTCA	AACAGGTAGT	ATCCTAGGTG	GAATTGAGTC	CAGCTCTGGC	3840
TCTGTAACAC	TTACTGCAAC	CGAGGGCGCT	CTTGCTGTAA	GCAATATTTC	GGGCAACACC	3900
GTTACTGTTA	CTGCAAATAG	CGGTGCATTA	ACCACTTTGG	CAGGCTCTAC	AATTAAAGGA	3960
ACCGAGAGTG	TAACCACCTTC	AAGTCAATCA	GGCGATATCG	GCGGTACGAT	TTCTGGTGGC	4020
ACAGTAGAGG	TTAAAGCAAC	CGAAAGTTA	ACCACTCAAT	CCAATTCAA	AATTAAAGCA	4080
ACAACAGGCG	AGGCTAACGT	AACAAGTGCA	ACAGGTACAA	TTGGTGGTAC	GATTTCGGT	4140
AATACGGTAA	ATGTTACGGC	AAACGCTGGC	GATTTAACAG	TTGGGAATGG	CGCAGAAATT	4200
AATGCGACAG	AAGGAGCTGC	AACCTTAACT	ACATCATCGG	GCAAATTAAAC	TACCGAAGCT	4260
AGTTCACACA	TTACTTCAGC	CAAGGGTCAG	GTAAATCTTT	CAGCTCAGGA	TGGTAGCGTT	4320
GCAGGAAGTA	TTAATGCCGC	CAATGTGACA	CTAAATACTA	CAGGCACCTT	AACTACCGTG	4380
AAGGGTTCAA	ACATTAATGC	AACCAGCGGT	ACCTTGGTTA	TTAACGCAA	AGACGCTGAG	4440
CTAAATGGCG	CAGCATTGGG	TAACCACACA	GTGGTAAATG	CAACCAACGC	AAATGGCTCC	4500
GGCAGCGTAA	TCGCGACAAAC	CTCAAGCAGA	GTGAACATCA	CTGGGGATT	AATCACAATA	4560
AATGGATTAA	ATATCATTTC	AAAAAACGGT	ATAAACACCG	TACTGTTAAA	AGGCGTTAAA	4620
ATTGATGTGA	AATACATTCA	ACCGGGTATA	GCAAGCGTAG	ATGAAGTAAT	TGAAGCGAAA	4680
CGCATCCTTG	AGAAGGTAAA	AGATTATCT	GATGAAGAAA	GAGAAGCGTT	AGCTAAACTT	4740
GGCGTAAGTG	CTGTACGTTT	TATTGAGCCA	AATAATACAA	TTACAGTCGA	TACACAAAAT	4800
GAATTGCAA	CCAGACCATT	AAGTCGAATA	GTGATTCTG	AAGGCAGGGC	GTGTTCTCA	4860
AACAGTGATG	GCGCGACGGT	GTGCGTTAAT	ATCGCTGATA	ACGGGCGGT	GCGGTAGTA	4920
ATTGACAAGG	TAGATTCAT	CCTGCAATGA	AGTCATT	TTTCGTATT	ATTTACTGTG	4980
TGGGTTAAAG	TTCAGTACGG	GCTTACCCA	TCTTGTAAAA	AATTACGGAG	AATACAATAA	5040
AGTATTTTA	ACAGGTTATT	ATTATGAAAA	ATATAAAAAG	CAGATTAAAA	CTCAGTGCAA	5100
TATCAGTATT	GCTTGGCCTG	GCTTCTTCAT	CATTGTATGC	AGAAGAAGCG	TTTTTAGTAA	5160
AAGGCTTCA	GTTATCTGGT	GCACCTGAAA	CTTTAAGTGA	AGACGCCAA	CTGTCTGTAG	5220
CAAAATCTTT	ATCTAAATAC	CAAGGCTCGC	AAACTTTAAC	AAACCTAAAA	ACAGCACAGC	5280
TTGAATTACA	GGCTGTGCTA	GATAAGATTG	AGCCAAATAA	GTTTGATGTG	ATATTGCCAC	5340
AAACAAACCAT	TACGGATGGC	AATATTATGT	TTGAGCTAGT	CTCGAAATCA	GCCGCAGAAA	5400
GCCAAGTTT	TTATAAGGGCG	AGCCAGGGTT	ATAGTGAAGA	AAATATCGCT	CGTAGCCTGC	5460
CATCTTGAA	ACAAGGAAAA	GTGTATGAAG	ATGGTCGTCA	GTGGTTCGAT	TTGCGTGAAT	5520

TCAATATGGC AAAAGAAAAT CCACTTAAAG TCACTCGCGT GCATTACGAG TTAAACCCCTA	5580
AAAACAAAAC CTCTGATTTG GTAGTTGCAG GTTTTCGCC TTTTGGCAAA ACGCGTAGCT	5640
TTGTTTCCTA TGATAATTTC GGCAGAAGGG AGTTAACTA TCAACGTGTA AGTCTAGGTT	5700
TTGTAATGC CAATTTGACC GGACATGATG ATGTATTAAA TCTAAACCCA TTGACCAATG	5760
TAAAAGCACC ATCAAAATCT TATGCGGTAG GCATAGGATA TACTTATCCG TTTTATGATA	5820
AACACCAATC CTTAAGTCTT TATACCAGCA TGAGTTATGC TGATTCTAAT GATATCGACG	5880
GCTTACCAAG TGCGATTAAT CGTAAATTAT CAAAAGGTCA ATCTATCTCT GCGAATCTGA	5940
AATGGAGTTA TTATCTCCCG ACATTTAACCC TTGGAATGGA AGACCAGTTT AAAATTAATT	6000
TAGGCTACAA CTACCGCCAT ATTAATCAA CATCCGAGTT AAACACCCCTG GGTGCAACGA	6060
AGAAAAAAATT TGCAGTATCA GGCAGTAAGTG CAGGCATTGA TGGACATATC CAATTTACCC	6120
CTAAAACAAT CTTTAATATT GATTTAACTC ATCATTATTA CGCGAGTAA TTACCCAGGCT	6180
CTTTTGGAAAT GGAGCGCATT GGCAGAACAT TTAATCGCAG CTATCACATT AGCACAGCCA	6240
GTTTAGGGTT GAGTCAAGAG TTTGCTCAAG GTTGGCATT TAGCAGTCAA TTATCGGGTC	6300
AGTTTACTCT ACAAGATATA AGTAGCATAG ATTTATTCTC TGTAACAGGT ACTTATGGCG	6360
TCAGAGGCTT TAAATACGGC GGTGCAAGTG GTGAGCGCGG TCTTGTATGG CGTAATGAAT	6420
TAAGTATGCC AAAATACACC CGCTTTCAAA TCAGCCCTTA TGCCTTTAT GATGCAGGTC	6480
AGTTCCGTTA TAATAGCGAA AATGCTAAA CTTACGGCGA AGATATGCAC ACGGTATCCT	6540
CTGCGGGTTT AGGCATTAAC ACCTCTCCTA CACAAAACCTT AAGCTTAGAT GCTTTGTTG	6600
CTCGTCGCTT TGCAAATGCC AATAGTGACA ATTTGAATGG CAACAAAAAA CGCACAAAGCT	6660
CACCTACAAC CTTCTGGGT AGATTAACAT TCAGTTCTA ACCCTGAAAT TTAATCAACT	6720
GGTAAGCGTT CCGCCTACCA GTTATATACT ATATGCTTTA CCCGCCAATT TACAGTCTAT	6780
ACGCAACCCCT GTTTTCATCC TTATATATCA AACAAACTAA GCAAACCAAG CAAACCAAGC	6840
AAACCAAGCA AACCAAGCAA ACCAAGCAA CCAAGCAAAC CAAGCAAACCA AAGCAAACCA	6900
AGCAAACCAA GCAAACCAAG CAAACCAAGC AAACCAAGCA ATGCTAAAAA ACAATTATA	6960
TGATAAACTA AAACATACTC CATAACCATGG CAATACAAGG GATTTAATAA TATGACAAAA	7020
GAAAATTTAC AAAGTGTTCAC AAAAATACG ACCGCTTCAC TTGTAGAATC AAACAACGAC	7080
CAAACCTCCC TGCAAATACT TAAACAACCA CCCAAACCCCA ACCTATTACG CCTGGAACCAA	7140
CATGTCGCCA AAAAAGATTA TGAGCTTGCT TGCGCGAAT TAATGGCGAT TTTGGAAAAAA	7200
ATGGACGCTA ATTTTGGAGG CGTTCACGAT ATTGAATTG ACGCACCTGC TCAGCTGGCA	7260
TATCTACCCG AAAAACTACT AATTCATTTC GCACTCGTC TCGCTAATGC AATTACAACA	7320
CTCTTTCCG ACCCCGAATT GGCAATTCC GAAGAAGGGG CATTAAAGAT GATTAGCCTG	7380
CAACGCTGGT TGACGCTGAT TTTGCTCT TCCCCCTACG TTAACGCAGA CCATATTCTC	7440
AATAAAATATA ATATCAACCC AGATTCCGAA GGTGGCTTTC ATTTAGCAAC AGACAACTCT	7500
TCTATTGCTA AATTCTGTAT TTTTACTTA CCCGAATCCA ATGTCAATAT GAGTTTAGAT	7560

GCCTTATGGG CAGGGAATCA ACAACTTGT GCTTCATTGT GTTTGCGTT GCAGTCTTCA	7620
CGTTTATTG GTACTGCATC TGCCTTCAT AAAAGAGCGG TGGTTTACA GTGGTTTCCT	7680
AAAAAACTCG CCGAAATTGC TAATTTAGAT GAATTGCCTG CAAATATCCT TCATGATGTA	7740
TATATGCACT GCAGTTATGA TTTAGCAAAA ACAAGCACG ATGTTAAGCG TCCATTAAAC	7800
GAACTTGTCC GCAAGCATAT CCTCACGCAA GGATGGCAAG ACCGCTACCT TTACACCTTA	7860
GGTAAAAAGG ACGGCAAACC TGTGATGATG GTACTGCTTG AACATTTAA TTCGGGACAT	7920
TCGATTTATC GCACGCATTC AACTCAATG ATTGCTGCTC GAGAAAAATT CTATTTAGTC	7980
GGCTTAGGCC ATGAGGGCGT TGATAACATA GGTCGAGAAG TGTTTGACGA GTTCTTGAA	8040
ATCAGTAGCA ATAATATAAT GGAGAGACTG TTTTTATCC GTAAACAGTG CGAAACTTTC	8100
CAACCCGCAG TGTTCTATAT GCCAAGCATT GGCAATGGATA TTACACGAT TTTTGTGAGC	8160
AACACTCGGC TTGCCCTAT TCAAGCTGTA GCCTTGGTC ATCCTGCCAC TACGCATTCT	8220
GAATTTATTG ATTATGTCA CGTAGAAGAT GATTATGTGG GCAGTGAAGA TTGTTTTAGC	8280
GAAACCTTT TACGCTTACC CAAAGATGCC CTACCTTATG TACCATCTGC ACTCGCCCCA	8340
AAAAAAAGTGG ATTATGTACT CAGGGAAAAC CCTGAAGTAG TCAATATCGG TATTGCCGCT	8400
ACCACAATGA AATTAAACCC TGAATTTTG CTAACATTGC AAGAAATCAG AGATAAAAGCT	8460
AAAGTCAAAA TACATTTCA TTTCGCACTT GGACAATCAA CAGGCTTGAC ACACCCCTAT	8520
GTCAAATGGT TTATCGAAAG CTATTTAGGT GACGATGCCA CTGCACATCC CCACGCACCT	8580
TATCAGCATT ATCTGGCAAT ATTGCGTGAT TGCGATATGC TACTAAATCC GTTCCCTTC	8640
GGTAATACTA ACGGCATAAT TGATATGGTT ACATTAGGTT TAGTTGGTGT ATGCAAAACG	8700
GGGGATGAAG TACATGAACA TATTGATGAA GGTCTGTTA AACGCTTAGG ACTACCAGAA	8760
TGGCTGATAG CCGACACACG AGAAACATAT ATTGAATGTG CTTTGCCTCT AGCAGAAAAC	8820
CATCAAGAAC GCCTTGAACT CCGTCGTTAC ATCATAGAAA ACAACGGCTT ACAAAAGCTT	8880
TTTACAGGCG ACCCTCGTCC ATTGGGAAA ATACTGCTTA AGAAAACAAA TGAATGGAAG	8940
CGGAAGCACT TGAGTAAAAA ATAACGGTTT TTTAAAGTAA AAGTGCCTT AATTTCAAA	9000
GCGTTTAAA AACCTCTCAA AAATCAACCG CACTTTATC TTTATAACGC TCCCGCGCGC	9060
TGACAGTTA TCTCTTCTT AAAATACCCA TAAAATTGTG GCAATAGTTG GGTAATCAAA	9120
TTCAATTGTT GATA CGGCAA ACTAAAGACG GCGCGTTCTT CGGCAGTCAT C	9171

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9323 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CGCCACTTCA	ATTTTGGATT	GTTGAAATTCA	AACTAACCAA	AAAGTCGGT	AAAAATCTGT	60
GGAGAAAATA	GGTTGTAGTG	AAGAACGAGG	TAATTGTTCA	AAAGGATAAA	GCTCTCTTAA	120
TTGGGCATTG	GTTGGCGTTT	CTTTTCGGT	TAATAGTAAA	TTATATTCTG	GACGACTATG	180
CAATCCACCA	ACAACCTTAC	CGTTGGTTTT	AAGCGTTAAT	GTAAGTTCTT	GCTCTCTTGG	240
GCGAATACGT	AATCCCATT	TTTGTGTTAGC	AAGAAAATGA	TCGGGATAAT	CATAATAGGT	300
GTTGCCAAA	AATAAATT	GATGTTCTAA	AATCATAAAAT	TTGCAAGAT	ATTGTGGCAA	360
TTCAATACCT	ATTTGTGGCG	AAATGCCAA	TTTTAATTCA	ATTTCTTGT	GCATAATATT	420
TCCCACCAA	ATCAACTGGT	TAAATATACA	AGATAATAAA	AATAAATCAA	GATTTTGTG	480
ATGACAAACA	ACAATTACAA	CACCTTTTT	GCAGTCTATA	TGCAAATATT	TTAAAAAAAT	540
AGTATAAAC	CGCCATATAA	AATGGTATAA	TCTTTCATCT	TTCATCTTTC	ATCTTTCATC	600
TTTCATCTT	CATCTTCAT	CTTTCATCTT	TCATCTTCA	TCTTTCATCT	TTCATCTTTC	660
ATCTTTCATC	TTTCATCTT	CACATGAAAT	GATGAACCGA	GGGAAGGGAG	GGAGGGGCAA	720
GAATGAAGAG	GGAGCTGAAC	GAACGCAAAT	GATAAAGTAA	TTTAATTGTT	CAACTAACCT	780
TAGGAGAAA	TATGAACAAG	ATATATCGTC	TCAAATTCA	CAAACGCCTG	AATGCTTTGG	840
TTGCTGTGTC	TGAATTGGCA	CGGGGTTGTG	ACCATTCCAC	AGAAAAAGGC	AGCGAAAAAC	900
CTGCTCGCAT	GAAAGTGC	GTGTTACATC	TTCCGCTATG	TTACTATCTT	960	
TAGGTGTAAC	ATCTATTCCA	CAATCTGTT	TAGCAAGCGG	CAATTAAACA	TCGACCAAAA	1020
TGAAATGGTG	CAGTTTTAC	AAGAAAACAA	GTAATAAAAC	CATTATCCGC	AACAGTGTG	1080
ACGCTATCAT	TAATTGGAAA	CAATTAAACA	TCGACCAAAA	TGAAATGGTG	CAGTTTTAC	1140
AAGAAAACAA	CAACTCCGCC	GTATTCAACC	GTGTTACATC	TAACCAAATC	TCCCAATTAA	1200
AAGGGATTTT	AGATTCTAAC	GGACAAGTCT	TTTTAATCAA	CCCAAATGGT	ATCACAATAG	1260
GTAAAGACGC	AATTATTAAC	ACTAATGGCT	TTACGGCTTC	TACGCTAGAC	ATTTCTAACG	1320
AAAACATCAA	GGCGCGTAAT	TTCACCTTCG	AGCAAACCAA	AGATAAAGCG	CTCGCTGAAA	1380
TTGTGAATCA	CGGTTAATT	ACTGTCGGTA	AAGACGGCAG	TGTAAATCTT	ATTGGTGGCA	1440
AAGTAAAAAA	CGAGGGTGTG	ATTAGCGTAA	ATGGTGGCAG	CATTTCTT	CTCGCAGGGC	1500
AAAAAATCAC	CATCAGCGAT	ATAATAAAC	CAACCATTAC	TTACAGCATT	GCCGCGCCTG	1560
AAAATGAAGC	GGTCAATCTG	GGCGATATT	TTGCCAAAGG	CGGTAAACATT	AATGTCGTG	1620
CTGCCACTAT	TCGAAACCAA	GGTAAACTT	CTGCTGATT	TGTAAGCAA	GATAAAGCG	1680
GCAATATTGT	TCTTCCGCC	AAAGAGGGTG	AAGCGGAAAT	TGGCGGTGTA	ATTCCGCTC	1740
AAAATCAGCA	AGCTAAAGGC	GGCAAGCTGA	TGATAAAAGTC	CGATAAAAGTC	ACATTAAAAA	1800
CAGGTGCAGT	TATCGACCTT	TCAGGTAAAG	AAGGGGGAGA	AACTTACCTT	GGCGGTGACG	1860
AGCGCGCGA	AGGTAAAAAC	GGCATTCAAT	TAGCAAAGAA	AACCTCTT	AAAAAAGGCT	1920
CAACCATCAA	TGTATCAGGC	AAAGAAAAAG	GC GGACGC	TATTGTGTGG	GGCGATATTG	1980

CGTTAATTGA CGGCAATATT AACGCTCAAG GTAGTGGTGA TATCGCTAAA ACCGGTGGTT	2040
TTGTGGAGAC ATCGGGGCAT TATTATCCA TTGACAGCAA TGCAATTGTT AAAACAAAAG	2100
AGTGGTTGCT AGACCCCTGAT GATGTAACAA TTGAAGCCGA AGACCCCCT CGCAATAATA	2160
CCGGTATAAA TGATGAATTG CCAACAGGCA CCGGTGAAGC AAGCGACCCCT AAAAAAAAATA	2220
GCGAACTCAA AACAAACGCTA ACCAATACAA CTATTCAAA TTATCTGAAA AACGCCTGGA	2280
CAATGAATAT AACGGCATCA AGAAAACCTTA CCGTTAATAG CTCAATCAAC ATCGGAAGCA	2340
ACTCCCACCTT AATTCTCCAT AGTAAAGGTC AGCGTGGCGG AGGCAGTCAG ATTGATGGAG	2400
ATATTACTTC TAAAGGCGGA AATTAAACCA TTTATTCTGG CGGATGGGTT GATGTTCATA	2460
AAAATATTAC GCTTGATCAG GGTTTTTAA ATATTACCGC CGCTTCCGTA GCTTTGAAG	2520
GTGAAATAA CAAAGCACGC GACGCGGCAA ATGCTAAAAT TGTGCCCCAG GGCACGTAA	2580
CCATTACAGG AGAGGGAAAA GATTTCAGGG CTAACAACGT ATCTTAAAC GGAACGGGTA	2640
AAGGTC TGAA TATCATTCA TCAGTGAATA ATTAAACCCA CAATCTTAGT GGCACAATTAA	2700
ACATATCTGG GAATATAACA ATTAACCAAA CTACGAGAAA GAACACCTCG TATTGGCAAA	2760
CCAGCCATGA TTCGCACTGG AACGTCAGTG CTCTTAATCT AGAGACAGGC GCAAATTTA	2820
CCTTTATTAA ATACATTCA AGCAATAGCA AAGGCTTAAC AACACAGTAT AGAAGCTCTG	2880
CAGGGGTGAA TTTAACGGC GTAAATGGCA ACATGTCATT CAATCTCAA GAAGGAGCGA	2940
AAGTTAATT CAAATTAAAA CCAAACGAGA ACATGAACAC AAGCAAACCT TTACCAATTG	3000
GGTTTTAGC CAATATCACA GCCACTGGTG GGGGCTCTGT TTTTTTGAT ATATATGCCA	3060
ACCATTCTGG CAGAGGGCT GAGTTAAAAA TGAGTGAAT TAATATCTCT AACGGCGCTA	3120
ATTTTACCTT AAATTCCCAT GTTCCGGCG ATGACGCTTT TAAAATCAAC AAAGACTTAA	3180
CCATAAATGC AACCAATTCA AATTTCAGCC TCAGACAGAC GAAAGATGAT TTTTATGACG	3240
GGTACGCCACG CAATGCCATC AATTCAACCT ACAACATATC CATTCTGGC GGTAAATGTCA	3300
CCCTTGGTGG ACAAAACTCA AGCAGCAGCA TTACGGGAA TATTACTATC GAGAAAGCAG	3360
CAAATGTTAC GCTAGAAGCC AATAACGCC CTAATCAGCA AAACATAAGG GATAGAGTTA	3420
TAAAACCTGG CAGCTTGCTC GTTAATGGGA GTTTAAGTTT AACTGGCGAA AATGCAGATA	3480
TTAAAGGCAA TCTCACTATT TCAGAAAGCG CCACTTTAA AGGAAAGACT AGAGATAACCC	3540
TAAATATCAC CGGCAATTAA ACCAATAATG GCACTGCCGA AATTAATATA ACACAAGGAG	3600
TGGTAAACT TGGCAATGTT ACCAATGATG GTGATTTAA CATTACCACT CACGCTAAAC	3660
GCAACCAAAG AAGCATCATC GGCAGGAGATA TAATCAACAA AAAAGGAAGC TTAAATATTA	3720
CAGACAGTAA TAATGATGCT GAAATCCAAA TTGGCGGCAA TATCTCGCAA AAAGAAGGCA	3780
ACCTCACGAT TTCTTCCGAT AAAATTAAATA TCACCAAACA GATAACAATC AAAAAGGGTA	3840
TTGATGGAGA GGACTCTAGT TCAGATGCGA CAAGTAATGC CAACCTAACT ATTAAAACCA	3900
AAGAATTGAA ATTGACAGAA GACCTAAGTA TTTCAGGTTT CAATAAGCA GAGATTACAG	3960
CCAAAGATGG TAGAGATTAA ACTATTGGCA ACAGTAATGA CGGTAACAGC GGTGCCGAAG	4020

CCAAAACAGT	AACTTTAAC	AATGTTAAG	ATTCAAAAT	CTCTGCTGAC	GGTCACAATG	4080
TGACACTAAA	TAGCAAAGTG	AAAACATCTA	GCAGCAATGG	CGGACGTGAA	AGCAATAGCG	4140
ACAACGATAAC	CGGCTTAAC	ATTACTGCAA	AAAATGTAGA	AGTAAACAAA	GATATTACTT	4200
CTCTAAAAC	AGTAAATATC	ACCGCGTCGG	AAAAGGTTAC	CACCACAGCA	GGCTCGACCA	4260
TTAACGCAAC	AAATGGCAA	GCAAGTATT	CAACCAAAAC	AGGTGATATC	AGCGGTACGA	4320
TTTCCGGTAA	CACGGTAAGT	GTTAGCGCGA	CTGGTGATTT	AACCAC	AA TCCGGCTCAA	4380
AAATTGAAGC	GAAATCGGGT	GAGGCTAATG	TAACAAGTGC	AACAGGTACA	ATTGGCGGTA	4440
CAATTTCGG	TAATACGGTA	AATGTTACGG	CAAACGCTGG	CGATTTAACA	GTTGGGAATG	4500
GCGCAGAAAT	TAATGCGACA	GAAGGAGCTG	CAACCTAAC	CGCAACAGGG	AATACCTTGA	4560
CTACTGAAGC	CGGTTCTAGC	ATCACTTCAA	CTAAGGGTCA	GGTAGACCTC	TTGGCTCAGA	4620
ATGGTAGCAT	CGCAGGAAGC	ATTAATGCTG	CTAATGTGAC	ATTAAATACT	ACAGGCACCT	4680
TAACCACCGT	GGCAGGCTCG	GATATTAAAG	CAACCAGCGG	CACCTTGGTT	ATTAACGCAA	4740
AAGATGCTAA	GCTAAATGGT	GATGCATCAG	GTGATAGTAC	AGAAGTGAAT	GCAGTCAACG	4800
ACTGGGGATT	TGGTAGTGTG	ACTGCGGCAA	CCTCAAGCAG	TGTGAATATC	ACTGGGGATT	4860
TAAACACAGT	AAATGGGTTA	AATATCATTT	CGAAAGATGG	TAGAAACACT	GTGCGCTTAA	4920
GAGGCAAGGA	AATTGAGGTG	AAATATATCC	AGCCAGGTGT	AGCAAGTGT	GAAGAAGTAA	4980
TTGAAGCGAA	ACCGTCCCTT	GAAAAAGTAA	AAGATTATAC	TGATGAAGAA	AGAGAAACAT	5040
TAGCTAAACT	TGGTGTAAAGT	GCTGTACGTT	TTGTTGAGCC	AAATAATACA	ATTACAGTCA	5100
ATACACAAAA	TGAATTTACA	ACCAAGACCGT	CAAGTCAAGT	GATAATTCT	GAAGGTAAGG	5160
CGTGTTCCTC	AAAGTGGTAAT	GGCGCACGAG	TATGTACCAA	TGTTGCTGAC	GATGGACAGC	5220
CGTAGTCAGT	AATTGACAAG	GTAGATTCA	TCCTGCAATG	AAGTCATTTT	ATTTTCGTAT	5280
TATTTACTGT	GTGGGTTAAA	GTTCAAGTACG	GGCTTACCC	ATCTTGTAAA	AAATTACGGA	5340
GAATACAATA	AAGTATTTT	AACAGGTTAT	TATTATGAAA	AATATAAAA	GCAGATTAAA	5400
ACTCAGTGCA	ATATCAGTAT	TGCTTGGCCT	GGCTTCTCA	TCATTGTATG	CAGAAGAAGC	5460
GTTTTAGTA	AAAGGCTTTC	AGTTATCTGG	TGCACTTGAA	ACTTTAAGTG	AAGACGCCA	5520
ACTGTCTGTA	GCAAAATCTT	TATCTAAATA	CCAAGGCTCG	CAAACTTAA	CAAACCTAAA	5580
AACAGCACAG	CTTGAATTAC	AGGCTGTGCT	AGATAAGATT	GAGCCAAATA	AATTTGATGT	5640
GATATTGCCG	CAACAAACCA	TTACGGATGG	CAATATCATG	TTTGAGCTAG	TCTCGAAATC	5700
AGCCGCAGAA	AGCCAAGTTT	TTTATAAGGC	GAGCCAGGGT	TATAGTGAAG	AAAATATCGC	5760
TCGTAGCCTG	CCATCTTGA	AACAAGGAAA	AGTGTATGAA	GATGGTCGTC	AGTGGTTCGA	5820
TTTGCCTGAA	TTTAATATGG	CAAAAGAAAA	CCCGCTTAAG	GTTACCCGTG	TACATTACGA	5880
ACTAAACCC	AAAAACAAA	CCTCTAATT	GATAATTGCG	GGCTTCTCGC	CTTTTGGTAA	5940
AACGCGTAGC	TTTATTTCTT	ATGATAATT	CGGCGCGAGA	GAGTTAACT	ACCAACGTGT	6000
AAGCTTGGGT	TTTGTAAATG	CCAATTAAAC	TGGTCATGAT	GATGTGTTAA	TTATACCAAGT	6060

ATGAGTTATG CTGATTCTAA TGATATCGAC GGCTTACCAA GTGCGATTAA TCGTAAATTA	6120
TCAAAAGGTC AATCTATCTC TCGAATCTG AAATGGAGTT ATTATCTCCC AACATTTAAC	6180
CTTGGCATGG AAGACCAATT TAAAATTAAT TTAGGCTACA ACTACCGCCA TATTAATCAA	6240
ACCTCCGCGT TAAATCGCTT GGGTGAAACG AAGAAAAAAT TTGAGTACATC AGGCGTAAGT	6300
GCAGGCATTG ATGGACATAT CCAATTTACC CCTAAAACAA TCTTTAATAT TGATTTAATC	6360
CATCATTATT ACGCGAGTAA ATTACCAGGC TCTTTGGAA TGGAGCGCAT TGGCGAAACA	6420
TTTAATCGCA GCTATCACAT TAGCACAGCC AGTTTAGGGT TGAGTCAAGA GTTTGCTCAA	6480
GGTTGGCATT TTAGCAGTCA ATTATCAGGT CAATTTACTC TACAAGATAT TAGCAGTATA	6540
GATTTATTCT CTGTAACAGG TACTTATGGC GTCAGAGGCT TAAATACGG CGGTGCAAGT	6600
GGTGAGCGCG GTCTTGTATG GCGTAATGAA TTAAGTATGC CAAAATACAC CCGCTTCCAA	6660
ATCAGCCCTT ATGCGTTTA TGATGCAGGT CAGTTCCGTT ATAATAGCGA AAATGCTAAA	6720
ACTTACGGCG AAGATATGCA CACGGTATCC TCTGCGGGTT TAGGCATTAA AACCTCTCCT	6780
ACACAAAAC TAAAGCTAGA TGCTTTGTT GCTCGTCGCT TTGCAAATGC CAATAGTGAC	6840
AATTTGAATG GCAACAAAAA ACGCACAAGC TCACCTACAA CCTTCTGGGG GAGATTAACA	6900
TTCAGTTCT AACCTGAAA TTTAATCAAC TGGTAAGCGT TCCGCCTACC AGTTTATAAC	6960
TATATGCTTT ACCCGCCAAT TTACAGTCTA TAGGCAACCC TGTTTTACC CTTATATATC	7020
AAATAAACAA GCTAAGCTGA GCTAAGCAAA CCAAGCAAC TCAAGCAAGC CAAGTAATAC	7080
TAAAAAAACA ATTTATATGA TAAACTAAAG TATACTCCAT GCCATGGCGA TACAAGGGAT	7140
TTAATAATAT GACAAAAGAA AATTTGCAAAC ACGCTCCTCA AGATGCGACC GCTTTACTTG	7200
CGGAATTAAG CAACAATCAA ACTCCCCTGC GAATATTTAA ACAACCACGC AAGCCAGCC	7260
TATTACGCTT GGAACAACAT ATCGAAAAAA AAGATTATGA GTTIGCTTGT CGTGAATTAA	7320
TGGTGATTCT GGAAAAAAATG GACGCTAATT TTGGAGGCAGT TCACGATATT GAATTTGACG	7380
CACCCGCTCA GCTGGCATAT CTACCCGAAA AATTACTAAT TTATTTGCC ACTCGTCTCG	7440
CTAATGCAAT TACAACACTC TTTCCGACC CCGAATTGGC AATTCTGAA GAAGGGCGT	7500
TAAAGATGAT TAGCCTGCAA CGCTGGTTGA CGCTGATTT TGCCCTTCC CCCTACGTTA	7560
ACGCAGACCA TATTCTCAAT AAATATAATA TCAACCCAGA TTCCGAAGGT GGCTTCATT	7620
TAGCAACAGA CAACTCTTCT ATTGCTAAAT TCTGTATTT TTACTTACCC GAATCCAATG	7680
TCAATATGAG TTTAGATGCG TTATGGCAG GGAATCAACA ACTTTGTGCT TCATTGTGTT	7740
TTGCGTTGCA GTCTTCACGT TTTATTGGTA CCGCATCTGC GTTTCATAAA AGAGCGGTGG	7800
TTTTACAGTG GTTTCCTAAA AAACCTGCCG AAATTGCTAA TTTAGATGAA TTGCCGTCAA	7860
ATATCCTTCA TGATGTATAT ATGCACTGCA GTTATGATTT AGCAAAAAAC AAGCACGATG	7920
TTAAGCGTCC ATTAAACGAA CTTGTCCGCA AGCATATCCT CACGCAAGGA TGGCAAGACC	7980
GCTACCTTCA CACCTTAGGT AAAAAGGACG GCAAACCTGT GATGATGGTA CTGCTTGAAC	8040
ATTTTAATTC GGGACATTGCG ATTTATCGTA CACATTCAAC TTCAATGATT GCTGCTCGAG	8100

AAAAAATTCTA	TTTAGTCGGC	TTAGGCCATG	AGGGCGTTGA	TAAAATAGGT	CGAGAAGTGT	8160
TTGACGAGTT	CTTTGAAATC	AGTAGCAATA	ATATAATGGA	GAGACTGTTT	TTTATCCGTA	8220
AACAGTGCAG	AACTTTCCAA	CCCGCAGTGT	TCTATATGCC	AAGCATTGGC	ATGGATATTA	8280
CCACGATTTT	TGTGAGCAAC	ACTCGGCTTG	CCCCTATTCA	AGCTGTAGCC	CTGGGTACATC	8340
CTGCCACTAC	GCATTCTGAA	TTTATTGATT	ATGTCATCGT	AGAAGATGAT	TATGTGGGCA	8400
GTGAAGATTG	TTTCAGCGAA	ACCCTTTAC	GCTTACCCAA	AGATGCCCTA	CCTTATGTAC	8460
CTTCTGCACT	CGCCCCACAA	AAAGTGGATT	ATGTACTCAG	GGAAAACCT	GAAGTAGTCA	8520
ATATCGGTAT	TGCCGCTACC	ACAATGAAAT	TAAACCCCTGA	ATTTTGCTA	ACATTGCAAG	8580
AAATCAGAGA	TAAAGCTAAA	GTCAAAATAC	ATTTTCATT	CGCACTTGGA	CAATCAACAG	8640
GCTTGACACA	CCCTTATGTC	AAATGGTTA	TCGAAAGCTA	TTTAGGTGAC	GATGCCACTG	8700
CACATCCCCA	CGCACCTTAT	CACGATTATC	TGGCAATATT	GCGTGATTGC	GATATGCTAC	8760
TAAATCCGTT	TCCTTTCGGT	AATACTAACG	GCATAATTGA	TATGGTTACA	TTAGGTTTAG	8820
TTGGGTGTATG	CAAAACGGGG	GATGAAGTAC	ATGAACATAT	TGATGAAGGT	CTGTTAAAC	8880
GCTTAGGACT	ACCAGAATGG	CTGATAGCCG	ACACACGAGA	AACATATATT	GAATGTGCTT	8940
TGCGTCTAGC	AGAAAACCAT	CAAGAACGCC	TTGAACCTCG	TCGTTACATC	ATAGAAAACA	9000
ACGGCTTACA	AAAGCTTTT	ACAGGCGACC	CTCGTCCATT	GGGCAAAATA	CTGCTTAAGA	9060
AAACAAATGA	ATGGAAGCGG	AAGCACTTGA	GTAAAAAATA	ACGGTTTTTT	AAAGTAAAAG	9120
TGCGGTTAAT	TTTCAAAGCG	TTTAAAAAC	CTCTAAAAAA	TCAACCGCAC	TTTTATCTTT	9180
ATAACGATCC	CGCACGCTGA	CAGTTTATCA	GCCTCCGCC	ATAAAACTCC	GCCTTTCATG	9240
GC GGAGATTT	TAGCCAAAAC	TGGCAGAAAT	TAAAGGCTAA	AATCACCAAA	TTGCACCACAA	9300
AAATCACCAA	TACCCACAAA	AAA				9323

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 4287 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GATCAATCTG	GGCGATATTT	TTGCCAAAGG	TGGTAACATT	AATGTCCGCG	CTGCCACTAT	60
TCGCAATAAA	GGTAAACTTT	CTGCCGACTC	TGTAAGCAAA	GATAAAAGTG	GTAACATTGT	120
TCTCTCTGCC	AAAGAAGGTG	AAGCGGAAAT	TGGCGGTGTA	ATTTCCGCTC	AAAATCAGCA	180
AGCCAAAGGT	GGTAAGTTGA	TGATTACAGG	CGATAAAAGTT	ACATTGAAAA	CGGGTGCACT	240
TATCGACCTT	TCGGGTAAAG	AAGGGGGAGA	AACTTATCTT	GGCGGTGACG	AGCGTGGCGA	300
AGGTAAAAAC	GGCATTCAAT	TAGCAAAGAA	AACCACTTTA	GAAAAAGGCT	CAACAATTAA	360

SUBSTITUTE SHEET (RULE 26)

TGTGTCAGGT AAAGAAAAAG CTGGGGCGCG TATTGTATGG GGCGATATTG CGTTAATTGA	420
CGGCAATATT AATGCCAAG GTAAAGATAT CGCTAAAACG GGTGGTTTG TGGAGACGTC	480
GGGGCATTAC TTATCCATTG ATGATAACGC AATTGTTAAA ACAAAAGAAT GGCTACTAGA	540
CCCAGAGAAT GTGACTATTG AAGCTCCTTC CGCTTCTCGC GTCGAGCTGG GTGCCGATAG	600
GAATTCCCAC TCGGCAGAGG TGATAAAAAGT GACCCTAAAA AAAAATAACA CCTCCTTGAC	660
AACACTAACCA AATACAACCA TTTCAAATCT TCTGAAAAGT GCCCACGTGG TGAAACATAAC	720
GGCAAGGAGA AAACCTTACCG TTAATAGCTC TATCAGTATA GAAAGAGGCT CCCACTTAAT	780
TCTCCACAGT GAAGGTCAGG GCGGTCAAGG TGTCAGATT GATAAAGATA TTACTTCTGA	840
AGGCAGGAAAT TTAACCATT ATTCTGGCGG ATGGGTTGAT GTTCATAAAA ATATTACGCT	900
TGGTAGCGGC TTTTTAAACA TCACAACTAA AGAAGGAGAT ATCGCCTTCG AAGACAAGTC	960
TGGACGGAAC AACCTAACCA TTACAGCCCA AGGGACCATC ACCTCAGGTA ATAGTAACGG	1020
CTTTAGATTT AACAAACGTCT CTCTAACAG CCTTGGCGGA AAGCTGAGCT TTACTGACAG	1080
CAGAGAGGAC AGAGGTTAGAA GAACTAAGGG TAATATCTCA AACAAATTG ACGGAACGTT	1140
AAACATTTCG GGAACGTAG ATATCTCAAT GAAAGCACCC AAAGTCAGCT GGTTTACAG	1200
AGACAAAGGA CGCACCTACT GGAACGTAAC CACTTTAAAT GTTACCTCGG GTAGTAAATT	1260
TAACCTCTCC ATTGACAGCA CAGGAAGTGG CTCAACAGGT CCAAGCATAAC GCAATGCAGA	1320
ATTAAATGGC ATAACATTAA ATAAAGCCAC TTTAATATC GCACAAGGCT CAACAGCTAA	1380
CTTTAGCATC AAGGCATCAA TAATGCCCTT TAAGAGTAAC GCTAACTACG CATTATTTAA	1440
TGAAGATATT TCAGTCTAG GGGGGGTAG CGTTAATTTC AAACCTAACG CCTCATCTAG	1500
CAACATACAA ACCCTGGCG TAATTATAAA ATCTAAAC TTTAATGTCT CAGGAGGGTC	1560
AACTTTAAAT CTCAAGGCTG AAGGTTCAAC AGAAAACGCT TTTTCAATAG AAAATGATTT	1620
AAACTAAAC GCCACCGGTG GCAATATAAC AATCAGACAA GTCGAGGGTA CCGATTACAG	1680
CGTCAACAAA GGTGTCGAG CCAAAAAAAA CATAACTTT AAAGGGGGTA ATATCACCTT	1740
CGGCTCTCAA AAAGCCACAA CAGAAATCAA AGGCAATGTT ACCATCAATA AAAACACTAA	1800
CGCTACTCTT CGTGGTGCAG ATTTGCCGA AAACAAATCG CCTTTAAATA TAGCAGGAAA	1860
TGTTATTAAT AATGGCAACC TTACCACTGC CGGCTCCATT ATCAATATAG CCGGAAATCT	1920
TACTGTTCA AAAGGCGCTA ACCTTCAGC TATAACAAAT TACACTTTA ATGTAGCCGG	1980
CTCATTGAC AACAAATGGCG CTTCAAACAT TTCCATTGCC AGAGGAGGGG CTAAATTTAA	2040
AGATATCAAT AACACCAAGTA GCTTAAATAT TACCACCAAC TCTGATACCA CTTACCGCAC	2100
CATTATAAAA GGCAATATAT CCAACAAATC AGGTGATTG AATATTATTG ATAAAAAAAAG	2160
CGACGCTGAA ATCCAAATTG GCGGCAATAT CTCACAAAAA GAAGGCAATC TCACAATTTC	2220
TTCTGATAAA GTAAATATTA CCAATCAGAT AACAAATCAA GCAGGCCTTG AAGGGGGCG	2280
TTCTGATTCA AGTGAGGCAG AAAATGCTAA CCTAACTATT CAAACCAAAG AGTTAAAATT	2340
GGCAGGAGAC CTAATATTT CAGGCTTAA TAAAGCAGAA ATTACAGCTA AAAATGGCAG	2400

TGATTTAACT ATTGGCAATG CTAGCGGTGG TAATGCTGAT GCTAAAAAAG TGACTTTGA	2460
CAAGGTTAAA GATTCAAAAA TCTCGACTGA CGGTACAAT GTAACACTAA ATAGCGAAGT	2520
GAAAACGTCT AATGGTAGTA GCAATGCTGG TAATGATAAC AGCACCGGTT TAACCATTTC	2580
CGCAAAAGAT GTAACGGTAA ACAATAACGT TACCTCCCAC AAGACAATAA ATATCTCTGC	2640
CGCAGCAGGA AATGTAACAA CCAAAGAAGG CACAACATAC AATGCAACCA CAGGCAGCGT	2700
GGAAGTAACT GCTCAAAATG GTACAATTAA AGGCAACATT ACCTCGAAA ATGTAACAGT	2760
GACAGCAACA GAAAATCTTG TTACCACAGA GAATGCTGTC ATTAATGCAA CCAGCGGCAC	2820
AGTAAACATT AGTACAAAAA CAGGGGATAT TAAAGGTGGA ATTGAATCAA CTTCCGGTAA	2880
TGTAAATATT ACAGCGAGCG GCAATACACT TAAGGTAAGT AATATCACTG GTCAAGATGT	2940
AACAGTAACA GCGGATGCAG GAGCCTTGAC AACTACAGCA GGCTCAACCA TTAGTGCAC	3000
AACAGGCAAT GCAAATATTA CAACCAAAAC AGGTGATATC AACGGTAAAG TTGAATCCAG	3060
CTCCGGCTCT GTAACACTTG TTGCAACTGG AGCAACTCTT GCTGTAGGTA ATATTCAGG	3120
TAACACTGTT ACTATTACTG CGGATAGCGG TAAATTAACC TCCACAGTAG GTTCTACAAT	3180
TAATGGGACT AATAGTGTAA CCACCTCAAG CCAATCAGGC GATATTGAAG GTACAATTTC	3240
TGGTAATACA GTAAATGTTA CAGCAAGCAC TGGTGATTAA ACTATTGGAA ATAGTGC	3300
AGTTGAAGCG AAAAATGGAG CTGCAACCTT AACTGCTGAA TCAGGCAAAT TAACCACCCA	3360
AACAGGCTCT AGCATTACCT CAAGCAATGG TCAGACAACCTT CTTACAGCCA AGGATAGCAG	3420
TATCGCAGGA AACATTAATG CTGCTAATGT GACGTTAAAT ACCACAGGCA CTTTAAC	3480
TACAGGGGAT TCAAAGATTA ACGCAACCAG TGGTACCTTA ACAATCAATG CAAAAGATGC	3540
CAAATTAGAT GGTGCTGCAT CAGGTGACCG CACAGTAGTA AATGCAACTA ACGCAAGTGG	3600
CTCTGGTAAC GTGACTGCGA AAACCTCAAG CAGCGTGAAT ATCACCGGGG ATTTAAACAC	3660
AATAAAATGGG TTAAATATCA TTTCGGAAAA TGGTAGAAAC ACTGTGCGCT TAAGAGGCAA	3720
GGAAATTGAT GTGAAATATA TCCAACCAGG TGTAGCAAGC GTAGAAGAGG TAATTGAAGC	3780
GAAACGCGTC CTTGAGAAGG TAAAAGATT ATCTGATGAA GAAAGAGAAA CACTAGCAA	3840
ACTTGGTGTA AGTGCTGTAC GTTTCGTTGA GCCAAATAAT GCCATTACGG TTAATACACA	3900
AAACGAGTTT ACAACCAAAAC CATCAAGTCA AGTGACAATT TCTGAAGGTA AGGCGTGT	3960
CTCAAGTGGT AATGGCGCAC GAGTATGTAC CAATGTTGCT GACGATGGAC AGCAGTAGTC	4020
AGTAATTGAC AAGGTAGATT TCATCCTGCA ATGAAGTCAT TTTATTTCG TATTATTTAC	4080
TGTGTGGGTT AAAGTTCACT ACGGGCTTTA CCCACCTTGT AAAAAATTAC GAAAATACA	4140
ATAAAGTATT TTTAACAGGT TATTATTATG AAAAACATAA AAAGCAGATT AAAACTCAGT	4200
GCAATATCAA TATTGCTTGG CTTGGCTTCT TCATCGACGT ATGCAGAAGA AGCGTTTTA	4260
GTAAAAGGCT TTCAGTTATC TGGCGCG	4287

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4702 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GGGAATGAGC GTCGTACACG GTACAGCAAC CATGCAAGTA GACGGCAATA AAACCACTAT	60
CCGTAATAGC ATCAATGCTA TCATCAATTG GAAACAATT AACATTGACC AAAATGAAAT	120
GGAGCAGTTT TTACAAGAAA GCAGCAACTC TGCCGTTTC AACCGTGTAA CATCTGACCA	180
AATCTCCCAA TTAAAAGGGA TTTTAGATTTC TAACGGACAA GTCTTTTAA TCAACCCAAA	240
TGGTATCACA ATAGGTAAAG ACGCAATTAT TAACACTAAT GGCTTTACTG CTTCTACGCT	300
AGACATTTCT AACGAAAACA TCAAGGCGCG TAATTTCACC CTTGAGCAAA CCAAGGATAAA	360
AGCACTCGCT GAAATCGTGA ATCACGGTTT AATTACCGTT GGTAAAGACG GTAGCGTAAA	420
CCTTATTGGT GGCAAAGTGA AAAACGAGGG CGTGATTAGC GTAAATGGCG GTAGTATTTC	480
TTTACTTGCA GGGCAAAAAA TCACCATCAG CGATATAATA AATCCAACCA TCACTTACAG	540
CATTGCTGCA CCTGAAAACG AAGCGATCAA TCTGGCGAT ATTTTGCCA AAGGTGGTAA	600
CATTAATGTC CGCGCTGCCA CTATTGCAA TAAAGGTAAA CTTTCTGCCG ACTCTGTAAG	660
CAAAGATAAA AGTGGTAACA TTGTTCTCTC TGCCAAAGAA GGTGAAGCGG AAATTGGCGG	720
TGTAATTTC GCTCAAAATC AGCAAGCCAA AGGTGGTAAG TTGATGATTA CAGGTGATAA	780
AGTCACATTA AAAACAGGTG CAGTTATCGA CCTTTCAGGT AAAGAAGGGG GAGAGACTTA	840
TCTTGGCGGT GATGAGCGTG GCGAAGGTAA AAATGGTATT CAATTAGCGA AGAAAACCTC	900
TTTAGAAAAA GGCTCGACAA TTAATGTATC AGGCAAAGAA AAAGGCAGGC GCGCTATTGT	960
ATGGGGCGAT ATTGCATTA TTAATGGTAA CATTAATGCT CAAGGTAGCG ATATTGCTAA	1020
AACTGGCGGC TTTGTGGAAA CATCAGGACA TGACTTATCC ATTGGTGATG ATGTGATTGT	1080
TGACGCTAAA GAGTGGTTAT TAGACCCAGA TGATGTGTCC ATTGAAACTC TTACATCTGG	1140
ACGCAATAAT ACCGGCGAAA ACCAAGGATA TACAACAGGA GATGGGACTA AAGAGTCACC	1200
TAAAGGTAAT AGTATTTCTA AACCTACATT AACAAACTCA ACTCTTGAGC AAATCCTAAG	1260
AAGAGGTTCT TATGTTAATA TCACTGCTAA TAATAGAATT TATGTTAATA GCTCCATCAA	1320
CTTATCTAAT GGCAGTTAA CACTTCACAC TAAACGAGAT GGAGTTAAA TTAACGGTGA	1380
TATTACCTCA AACGAAAATG GTAATTAAAC CATTAAAGCA GGCTCTTGGG TTGATGTTCA	1440
TAAAAACATC ACGCTTGGTA CGGGTTTTT CAATATTGTC GCTGGGGATT CTGTAGCTTT	1500
TGAGAGAGAG GGCGATAAAAG CACGTAACGC AACAGATGCT CAAATTACCG CACAAGGGAC	1560
GATAACCGTC AATAAAGATG ATAAACAATT TAGATTCAAT AATGTATCTA TTAACGGGAC	1620

SUBSTITUTE SHEET (RULE 26)

GGGCAAGGGT TTAAAGTTA TTGCAAATCA AAATAATTTC ACTCATAAAAT TTGATGGCGA	1680
AATTAACATA TCTGGAATAG TAACAATTAA CCAAACCACG AAAAAAGATG TTAAATACTG	1740
GAATGCATCA AAAGACTCTT ACTGGAATGT TTCTTCTCTT ACTTTGAATA CGGTGCAAA	1800
ATTTACCTT ATAAAATTG TTGATAGCGG CTCAAATTCC CAAGATTGA GGTACATCAG	1860
TAGAAGTTT GCAGGCGTAC ATTTAACGG CATCGGAGGC AAAACAAACT TCAACATCGG	1920
AGCTAACGCA AAAGCCTTAT TTAAATTAAA ACCAAACGCC GCTACAGACC CAAAAAAAGA	1980
ATTACCTATT ACTTTAACG CCAACATTAC AGCTACCGGT AACAGTGATA GCTCTGTGAT	2040
GTTTGACATA CACGCCAATC TTACCTCTAG AGCTGCCGGC ATAAACATGG ATTCAATTAA	2100
CATTACCGGC GGGCTTGACT TTTCCATAAC ATCCCATAAT CGCAATAGTA ATGCTTTGA	2160
AATCAAAAAA GACTTAACTA TAAATGCAAC TGGCTCGAAT TTTAGTCTTA AGCAAACGAA	2220
AGATTCTTT TATAATGAAT ACAGCAAACA CGCCATTAAC TCAAGTCATA ATCTAACCAT	2280
TCTTGGCGGC AATGTCACTC TAGGTGGGG AATTCAGC AGTAGCATT CGGGCAATAT	2340
CAATATCACC AATAAAGCAA ATGTTACATT ACAAGCTGAC ACCAGCAACA GCAACACAGG	2400
CTTGAAGAAA AGAACTCTAA CTCTTGGCAA TATATCTGTT GAGGGGAATT TAAGCCTAAC	2460
TGGTGCAAAT GCAAACATTG TCGGCAATCT TTCTATTGCA GAAGATTCCA CATTAAAGG	2520
AGAAGCCAGT GACAACCTAA ACATCACCGG CACCTTTACC AACAAACGGTA CGGCCAACAT	2580
TAATATAAAA CAAGGAGTGG TAAAACCTCA AGGCGATATT ATCAATAAAG GTGGTTAAA	2640
TATCACTACT AACGCCTCAG GCACTAAAA AACCATTATT AACGGAAATA TAACTAACGA	2700
AAAAGGCGAC TTAAACATCA AGAATATTAA AGCCGACGCC GAAATCCAAA TTGGCGCAA	2760
TATCTCACAA AAAGAAGGCA ATCTCACAAT TTCTTCTGAT AAAGTAAATA TTACCAATCA	2820
GATAACAATC AAAGCAGGCG TTGAAGGGGG GCGTTCTGAT TCAAGTGAGG CAGAAATGC	2880
TAACCTAACT ATTCAAACCA AAGAGTTAAA ATTGGCAGGA GACCTAAATA TTTCAGGCTT	2940
TAATAAAGCA GAAATTACAG CTAAAATGG CAGTGATTAA ACTATTGGCA ATGCTAGCGG	3000
TGGTAATGCT GATGCTAAA AAGTGACTTT TGACAAGGTT AAAGATTCAA AAATCTCGAC	3060
TGACGGTCAC AATGTAACAC TAAATAGCGA AGTAAAACG TCTAATGGTA GTAGCAATGC	3120
TGGTAATGAT AACAGCACCG GTTTAACCAT TTCCGAAAAA GATGTAACGG TAAACAATAA	3180
CGTTACCTCC CACAAGACAA TAAATATCTC TGCCGCAGCA GGAAATGTAA CAACCAAAGA	3240
AGGCACAAC ATCAATGCAA CCACAGGCAG CGTGGAAAGTA ACTGCTAAA ATGGTACAAT	3300
TAAAGGCAAC ATTACCTCGC AAAATGTAAC AGTGACAGCA ACAGAAAATC TTGTTACCAC	3360
AGAGAAATGCT GTCATTAATG CAACCAGCGG CACAGTAAAC ATTAGTACAA AAACAGGGGA	3420
TATTAAGGT GGAATTGAAT CAACTTCCGG TAATGTAAAT ATTACAGCGA GCGGCAATAC	3480
ACTTAAGGTA AGTAATATCA CTGGTCAAGA TGTAACAGTA ACAGCGGATG CAGGAGCCTT	3540
GACAACATACA GCAGGCTCAA CCATTAGTGC GACAACAGGC AATGCAAATA TTACAACCAA	3600
AACAGGTGAT ATCAACGGTA AAGTTGAATC CAGCTCCGGC TCTGTAACAC TTGTTGCAAC	3660

TGGAGCAACT CTTGCTGTAG	GTAATATTC AGGTAACACT	GTTACTATTA CTGGGATAG	3720
CGGTAAATTA ACCTCCACAG	TAGGTTCTAC AATTAATGGG	ACTAATAGTG TAACCACCTC	3780
AAGCCAATCA GGCGATATTG	AAGGTACAAT TTCTGGTAAT	ACAGTAAATG TTACAGCAAG	3840
CACTGGTGAT TTAACTATTG	GAAATAGTGC AAAAGTTGAA	GCGAAAAATG GAGCTGCAAC	3900
CTTAACTGCT GAATCAGGCA	AATTAACCAC CCAAACAGGC	TCTAGCATT CCTCAAGCAA	3960
TGGTCAGACA ACTCTTACAG	CCAAGGATAG CAGTATCGCA	GGAAACATTA ATGCTGCTAA	4020
TGTGACGTTA AATACCACAG	GCACTTAAC TACTACAGGG	GATTCAAAGA TTAACGCAAC	4080
CAGTGGTACC TTAACAATCA	ATGCAAAAGA TGCCAAATTA	GATGGTGCTG CATCAGGTGA	4140
CCGCACAGTA GTAAATGCAA	CTAACGCAAG TGGCTCTGGT	AACGTGACTG CGAAAACCTC	4200
AAGCAGCGTG AATATCACCG	GGGATTAAA CACAATAAAT	GGGTTAAATA TCATTCGGA	4260
AAATGGTAGA AACACTGTGC	GCTTAAGAGG CAAGGAAATT	GATGTGAAAT ATATCCAACC	4320
AGGTGTAGCA AGCGTAGAAG	AGGTAAATTGA AGCGAACGC	GTCCTTGAGA AGGTAAAAGA	4380
TTTATCTGAT GAAGAAAGAG	AAACACTAGC CAAACTTGGT	GTAAGTGCTG TACGTTTCGT	4440
TGAGCCAAAT AATGCCATTA	CGGTTAATAC ACAAAACGAG	TTTACAACCA AACCATCAAG	4500
TCAAGTGACA ATTTCTGAAG	GTAAGGCGTG TTTCTCAAGT	GGTAATGGCG CACGAGTATG	4560
TACCAATGTT GCTGACGATG	GACAGCAGTA GTCAGTAATT	GACAAGGTAG ATTCATCCT	4620
GCAATGAAGT CATTATTTT	TCGTATTATT TACTGTGTGG	GTAAAGTTC AGTACGGGCT	4680
TTACCCACCT TGTAAAAAT TA			4702

CLAIMS

What we claim is:

1. A vaccine against disease caused by non-typeable Haemophilus influenzae, including otitis media, sinusitis and bronchitis, comprising an effective amount of a high molecular weight protein of non-typeable Haemophilus influenzae which is protein HMW1, HMW2, HMW3 or HMW4 or a variant or fragment of said protein retaining immunological properties thereof or a synthetic peptide having an amino acid sequence corresponding to that of said protein, and a physiological carrier therefor.
2. The vaccine of claim 1 wherein said protein is HMW1 encoded by the DNA sequence shown in Figure 1 (SEQ ID NO:1), having the derived amino acid sequence of Figure 2 (SEQ ID NO:2) and having an apparent molecular weight of 125 kDa.
3. The vaccine of claim 1 wherein said protein is HMW2 encoding by the DNA sequence shown in Figure 3 (SEQ ID NO:3), having the derived amino acid sequence of Figure 4 (SEQ ID NO:4) and having an apparent molecular weight of 120 kDa.

SUBSTITUTE SHEET (RULE 26)

FIG. 1A. DNA SEQUENCE OF HIGH MOLECULAR WEIGHT PROTEIN

I (HMW1)

1 ACAGGGTTCTT CTTAATACCA GTACAAACCC ACAATAAAAT ATGACAAACA
 51 ACAATTACAA CACCTTTTTT GCAGCTATA TGCAAATATT TTAAAAAATA
 101 GTATAATCC GCCATATAAA ATGGTATAAT CTTCATCTT TCATCTTTCA
 151 TCTTTCATCTT TTICATCTTTC ATCTTTCATC TTTCATCTT CATCTTTCAT
 201 CTTTCATCTT TCATCTTCA TCTTTCATCTT TTCACTCTTTC ACATGCCCTG
 251 ATGAAACGGAG GGAAGGGAGG GAGGGCAAG AATGAAGAGG GAGCTGAACG
 301 AACGCAAATG ATAAAGTAAT TTAATTTGTTT AACCTAACCTT AGGAGAAAT
 351 ATGAAACAAAGC TATATCGTCT CAAATTCAGC AAACGCCTGA ATGCTTTGGT
 401 TGCTGTGTCT GAATTGGCAC GGGGTGTGA CCATTCCACCA GAAAAGGCA
 451 GCGAAAAACC TGCTCGCATG AAAGTGGTC ACTTTAGGGTT AAAGCCACTT
 501 TCCGCTATGTT TACTATCTT AGGTGTAACA TCTATTCCAC AATCTGTTT
 551 AGCAAGGGC TTACAAGGAA TGGATGTAGT ACACGGCACA GCCACTATGC
 601 AAGTAGATGG TAATAAAACC ATTATCCGA ACAGTGTGTA CGATATCATT
 651 AATTGGAAAC AATTAAACAT CGACCAAAAT GAAATGGTGC AGTTTTTACA
 701 AGAAAACAAAC AACTCCGCCG TATTCAACCG TGTACATCT AACCAAATCT

FIG. 1B.

751 CCCAATTAAA AGGGATTTA GATTCTAACG GACAAGTCTT TTTAATCAAC
 801 CCAAATGGTA TCACAAATTAGG TAAAGACGCA ATTATTAAACA CTAATGGCTT
 851 TACGGCTTCT ACGGCTAGACA TTTCTAACGA AAACATCAAG GCGCGTAATT
 901 TCACCTTCGA GCAAACCAA GATAAAGCGC TCGCTGAAAT TGTGAATCAC
 951 GTTTAATT CTGTCGGTAA AGACGGCAGT GTAAATCTTA TTGGTGGCAA
 1001 AGTGAACAC GAGGGTGTGA TTAGCGTAAA TGGTGGCAGC ATTTCCTTTAC
 1051 TCGCAGGGCA AAAAATCACC ATCAGCGATA TAATAAACCC ACCATTACT
 1101 TACAGCATTG CCGCGCTGA AAATGAAGCG GTCAATCTGG GCGATATT
 1151 TGCCAAAGGC GGTAAACATTA ATGTCCCGTGC TGCCACTATT CGAAACCAAG
 1201 GTAAAACTTTC TGCTGATTTCT GTAAAGCAAAG ATAAAAGCGG CAATATTGTT
 1251 CTTTCCGCCA AAGAGGGTGA AGCGGAAATT GGGGGTGTAA TTTCCGCTCA
 1301 AAATCAGCAA GCTAAAGGCG GCAAGCTGAT GATTACAGGC GATAAAGTCA
 1351 CATTAAAC AGGTGCAGCTT ATCGACCTTT CAGGTAAGA AGGGGGAGAA
 1401 ACTTACCTTG GCGGTGACGA GCGCGGGCGAA GGTAAGGG GCATTCAAATT
 1451 AGCAAAGAAA ACCTCTTTAG AAAAGGCTC AACCATCAAT GTATCAGGCA
 1501 AAGAAAAGG CGGACGGCGCT ATTGTGTGGG GCGGATATTGC GTTAATTGAC

2 / 6

FIG. 1C.

1551 GGCATAATTACCGCTCAAGG TAGTGGTGAT ATCGCTAAAA CGGGTGGTT
 1601 TGTGGAGACG TCGGGCATG ATTTATTCTAT CAAAGACAAAT GCAATTGTTG
 1651 ACGCCAAGA GTGGTTGTTA GACCCGGATA ATGTTATCTAT TAATGCAGAA
 1701 ACAGCAGGAC GCAGCAATAC TTTCAGAAGAC GATGAATAACA CGGGATCCGG
 1751 GAATAGTGCC AGCACCCAA AACGAAACAA AGAAAAGACA ACATTAACAA
 1801 ACACAACTCT TGAGAGTATA CTAAAAAAAG GTACCTTTGT TAACATCACT
 1851 GCTAATCAC GCATCTATGT CAATAGCTCC ATTAAATTAT CCAATGGCAG 3 / 68
 1901 CTTAACTCTT TGAGTGAGG GTCGGAGGGG TGGGGGGTT GAGATTAACA
 1951 ACGATATTAC CACCGGTGAT GATACCAGAG GTGCAAACCTT ACAATTTAC
 2001 TCAGGGGCT GGGTTGATGT TCATAAAAAT ATCTCACTCG GGGCGCAAGG
 2051 TAACATAAAC ATTACAGCTA AACAAAGATAT CGCCTTTGAG AAAGGAAGCA
 2101 ACCAAGTCAT TACAGGTCAA GGGACTATTA CCTCAGGCAA TCAAAAGGT
 2151 TTAGATTAA ATAATGTCCTC TCTAAACGGC ACTGGCAGCG GACTGCAATT
 2201 CACCACTAA AGAACCAATA AATAACGCTAT CACAAATAAA TTTGAAGGGA
 2251 CTTAAATAT TTCAAGGAAA GTGAACATCT CAATGGTTTT ACCTAAAAAT
 2301 GAAAGTGGAT ATGATAAATT CAAAGGACGC ACTTACTGGA ATTAAACCTC

FIG. 1D.

2351 CTTAAATGTT TCCGAGAGTG GCGAGTTAA CCTCACTATT GACTCCAGAG
 2401 GAAGCGATAG TGCAGGCACA CTTACCCAGC CTTATAATT AAACGGTATA
 2451 TCATTCACCA AAGACACTAC CTTTAATGTT GAACGAAATG CAAGAGTCAA
 2501 CTTTGACATC AAGGCACCAA TAGGGATAAA TAAGTATTCT AGTTTGATT
 2551 ACGCATCAT TAAATGAAAC ATTTCACTT CGGGAGGGG GAGTGTGAT
 2601 TTCACACTTC TCGCCTCATC CTCTAACGTC CAAACCCCCG GTGTAGTTAT
 2651 AAATTCCTAA TACTTTAATG TTTCACAGG GTCAAGTTA AGATTTAAA 4/60
 2701 CTTCAAGGCTC AACAAAAACT GGCTTCTCAA TAGAGAAAGA TTTAACTTTA
 2751 AATGCCACCG GAGGCCAACAT AACACTTTG CAAGTTGAAG GCACCGATGG
 2801 AATGATTGGT AAGGCATTG TAGCCAAAA AACATAACC TTTGAAGGAG
 2851 GTAACATCAC CTTTGGCTCC AGGAAAGCCG TAACAGAAAT CGAAGGCAAT
 2901 GTTACTATCA ATAACAACGC TAACGTCACT CTTATCGGTT CGGATTGTGA
 2951 CAACCATCAA AACCTTTAA CTATTTAA AGATGTCACT ATTAAATAGCC
 3001 GCAACCTTAC CGCTGGAGGC ATATTTGTCA ATATAGCCGG AAATCTTAC
 3051 GTTGAAAGTA ACGCTAATT CAAAGCTATC ACAAAATTCA CTTTTAATGT
 3101 AGGGGGCTTG TTGACACAA AGGCAATT AAATATTCC ATTGCCAAG
 3151 GAGGGGCTCG CTTAAAGAC ATTGATAATT CCAAGAATT AAGCATCACC

FIG. 1E.

3201 ACCAACTCCA GCTCCACTTA CCGCACTATT ATAAGCGGCA ATATAACCAA
 3251 TAAAACGGT GATTAAATA TTACGGAACGA AGGTAGTGAT ACTGAAATGC
 3301 AAATTGGGG CGATGTCTCG CAAAAGAACG GTAATCTCAC GATTCTTCT
 3351 GACAAATCA ATATTACCAA ACAGATAACA ATCAAGGCAG GTGTTGATGG
 3401 GGAGAATTCC GATTCAAGCG CGACAAACAA TGCCAATCTA ACCATTTAAA
 3451 CCAAAGAATT GAAATTAAACG CAAGACCTAA ATATTCAAGG TTTCAATAAA
 3501 GCAGGAGATA CAGCTAAAGA TGGTAGTGAT TTAACTATTG GTAACACCAA 5/68
 3551 TAGTGGCTGAT GGTACTAATG CCAAAAAGT AACCTTTAAC CAGGTTAAAG
 3601 ATTCAAAAT CTCTGCTGAC GGTCAACAAGG TGACACTACA CAGCAAAGTG
 3651 GAAACATCCG GTAGTAATAA CAACACTGAA GATAGCAGTG ACAATAATGC
 3701 CGGCTTAACT ATCGATGCAA AAAATGTAAC AGTAAACAAAC ATATTTACTT
 3751 CTCACAAAGC AGTGAGGCATC TCTGGCACAA GTGGAGAAAT TACCACTAAA
 3801 ACAGGTACAA CCATTAACGC AACCACTGGT AACGTGGAGA TAACCGCTCA
 3851 AACAGGTAGT ATCCTAGGTG GAATTGAGTC CAGCTCTGGC TCTGTTAACAC
 3901 TTACTGCAAC CGAGGGCGCT CTTGCTGTAA GCAATATTTC GGGCAACACC
 3951 GTTACTGTAA CTGCAAATAG CGGTGCATTA ACCACTTTGG CAGGCTCTAC

FIG. 1F.

4001 AATTAAAGGA ACCGAGAGTG TAACCACTTC AAGTCATCA GGGGATATCG
 4051 GCGGTACGAT TTCTGGTGGC ACAGTAGAGG TAAAGCAAC CGAAAGTTA
 4101 ACCACTCAAT CCAATTCAA AATTAAAGCA ACAACAGGG AGGCTAACGT
 4151 AACAAAGTGCA ACAGGTACAA TTGGTGGTAC GATTTCGGT AATACGGTAA
 4201 ATGTTACGGC AAACGGCTGGC GATTAAACAG TTGGGATGG CGCAGAAATT
 4251 AATGCCACAG AAGGAGCTGC AACCTTAAC ACTCATCGG GCAAATTAAC
 4301 TACCGAAGCT AGTTACACACA TTACTTCAGC CAAGGGTCAG GTAAATCTTT
 4351 CAGCTCAGGA TGGTAGCGGT GCAGGAAGTA TTAATGCCGC CAATGTGACA
 4401 CTAATACTA CAGGCACTT AACTACCGTG AAGGGTTCAA ACATTAAATGC
 4451 ACCCAGGGT ACCTTGGTAA TTAACGCAA AGACGGCTGAG CTAATGGCG
 4501 CAGCATGGG TAACCACACA GTGGTAAATG CAAACCAACGC AAATGGCTCC
 4551 GGCAGGGTAA TCGCCGACAAC CTCAGCAGA GTGAACATCA CTGGGGATT
 4601 AATCACAAATA AATGGATTAA ATATCATTTC AAAAACGGT ATAAACACCG
 4651 TACTGTAAA AGGGCGTTAAA ATTGATGTGA AATACATTCAC ACCGGGTATA
 4701 GCAAGGGTAG ATGAAGTAAT TGAAGCGAAA CGCATCCTTG AGAAGGTAAA
 4751 AGATTTATCT GATGAAGAAA GAGAAGCGTT AGCTAAACTT GGAGTAAAGTG
 4801 CTGTACGTTT TATTGAGCCA AATAATACAA TTACAGTCGA TACACAAAT

6/68

7/68

FIG. 1G.

4851	GAATTGCAA	CCAGACCATT	AAGTCGAATA	GTGATTCTTG	AAGGCAGGGC
4901	GTGTTCTCA	AACAGTGATG	GCCGACGGT	GTGCCGTTAAT	ATCGCTGATA
4951	ACGGCGGTA	GCGGTCACTA	ATTGACAAAGG	TAGATTTCAT	CCTGCAATGA
5001	AGTCATTAA	TTTTCGTATT	ATTACTGTG	TGGGTTAAAG	TCAGTACGG
5051	GCTTTACCCA	TCTTGTAATA	AATTACGGAG	AATACAATAA	AGTATTTTA
5101	ACAGGTTATT	ATTATG			

FIG. 2A. AMINO ACID SEQUENCE OF HIGH MOLECULAR WEIGHT**PROTEIN I**

1 MNKIVRLKFS KRLNALVAVS ELARGCDHST EKGSEKPARM KVRHLALKPL
 51 SAMLSSLGVVT SIPQSVLASS LQGMDVVHGT ATMQVDGMKT IIRNSVDAII
 101 NWKQFNIDQN EMVQFLQENN NSAVFNRVTS NQISQLKGIL DSNGQVFLIN
 151 PNGITIIGKDA IINTNGFTAS TLDISNENIK ARNFTFEQTK DKALAEIVNH
 201 GLITVGKDGS VNLLIGGKVKN EGVISVNGGS ISLLAGQKIT ISDIINPTIT
 251 YSIAAPENEAV VNLGDIIFAKG GNINVRRAATI RNQGKLSADS VSKDKSGNIV
 301 LSAKEGEAEI GGVISAQNQQ AKGGKILMITG DVVTLLKTGAV IDLSGREGGE
 351 TYLGGDERGE GKNGIQLAKK TSLEKGSTIN VSGKEKGGRA IVWGDIALLID
 401 GNINAQGSGD IAKTGGFVET SGHDLFIKDN AIVDAKEWLL DFDNVSINAE
 451 TAGRSNTSED DEYTGGNSA STPKRNRKEKT TLTTNTLESI LKKGTFVNIT
 501 ANQRIVVNSS INLSNGSLTL WSEGRSGGGV EINNDITTGD DTRGANLTIY
 551 SGGWVDVHKN ISLGAQGNIN ITAKQDIAFE KGSNQVITGQ GTITSGNQKG
 601 FRFMIVSLLNG TGSGLQFTTK RTNKYAITNK FEGTLLNISGK VNISMVLPKN
 651 ESGYDKFKGR TYWNLTSLNV SESGEFNLTI DSRGSDSAGT LTQPYNLNGI
 701 SFNKDTTFNV ERNARVNFDI KAPIGINKYS SLNYASFNGN ISVSGGGGSVD

80 / 60

FIG. 2B.

751 FTLLASSSNV QTPGVVVINSK YFNVSTGSSL RFKTSGSTKT GFSIEKDLTL
 801 NATGCNITLL QVEGTDGMIG KGIVAKKKNIT FEGGNITFGS RKAUTIEGN
 851 VTINNNANVT LIGSDFDNHQ KPLTIKKDVT INSGNLTAGG NIVNIAGNLT
 901 VESNANFKAI TNFTFNVGGL FDNKGNNSNIS IAKGGARFKD IDNSKNLNSIT
 951 TNSSSTYRTI ISGNINITKNG DLNITNEGSD TEMQIGGGDVS QKEGNLTISS
 1001 DKINITKQIT IKAGVDGENS DSDATNNANL TIKTKELKLT QDLNISGFNK
 1051 AEITAKDGSD LTIGNNTNSAD GTNAKKVTFN QVKDKDSKISAD GHKVTLHSKV 9/68
 1101 ETSGSNNNTE DSSDNNNAGLT IDAKNVTVNN NITSHKAVSI SATSGEITTK
 1151 TGTTINATTG NVEITAQTGS ILGGIESSSG SVLTATEGA LAVSNISGNT
 1201 VTVVTANSGAL TTLAGSTIKG TESVTTSSQS GDIGGTTISGG TVEVKATESL
 1251 TTQNSNSKIKA TTGEANVTSA TGTIGGTISG NTVNVNTANAG DLTVNGNGAEI
 1301 NATEGAATLT TSSGKLTEA SSHITSAKGQ VNLSAQDGSV AGSINAANVT
 1351 LNTTGTLLTV KGSNNINATSG TLVINAKDAE LNGAALGNHT VVNATNANGS
 1401 GSVIATTSSR VNITGDLITI NGLNNIISKNG INTVLLKGVK IDVKYIOPGI
 1451 ASVDEVIEAK RILEKVKDLS DEEREALAKL GVSAVRFIEP NNTITVDTQN
 1501 EFATRPLSRI VISEGRACFS NSDGATVCVN IADNGR

FIG. 3A. AMINO ACID SEQUENCE OF HIGH MOLECULAR WEIGHT PROTEIN II (HMW2)

1 TAAATACA AGATAATAAA AATAAAATCAA GATTTTTGTG ATGACAAACA
 51 ACAATTACAA CACCTTTT GCAGTCTATA TGCAAATATT TTAAAAAAAT
 101 AGTATAAATC CGCCATATAA AATGGTATAA TCTTCATCT TTCACTTTA
 151 ATCTTCATC TTTCATCTTT CATCTTCAT CTTTCATCTT TCATCTTCA
 201 TCTTTCATCT TTTCATCTTC ATCTTCATC ATCTTCATCTT CACATGAAAT
 251 GATGAAACCGA GGGAAAGGGAG GGAGGGGCAA GAATGAAGAG GGAGCTGAAC 10 /
 301 GAACGCCAAT GATAAAAGTAA TTTAATTGTT CAACTAACCT TAGGAGAAA 68
 351 TATGAAACAAAG ATATATGTC TCAAATTCAAG CAAACGCCCTG AATGCTTTGG
 401 TTGGCTGTGTC TGAATTGGCA CGGGGTTGTG ACCATTCCAC AGAAAAAGGC
 451 TTCCGCTATG TTACTATCTT TAGGTGTAAAC CACTTAGCGT TAAAGCCACT
 501 TTCCGCTATG TTACTATCTT TAGGTGTAAAC ATCTATTCCA CAATCTGTGTT
 551 TAGCAAGCGG CTTACAAAGGA ATGGATGTAG TACACGGCAC AGCCACTATG
 601 CAAGTAGATG GTAAATAAAC CATTATCCGC AACAGTGTG ACGCTATCAT
 651 TAATTGGAAA CAATTAAACA TCGACCAAA TGAAATGGTG CAGTTTTAC
 701 AAGAAACAA CAACTCCGCC GTATTCAACC GTGTTACATC TAACCAAATC

FIG. 3B.

751 TCCCAATTAA AAGGGATT AGATTCTAAC GGACAAGTCT TTTTAAATCAA
 801 CCCAAATGGT ATCACAAATAG GTAAAGACGC AATTATTAAC ACTAATGGCT
 851 TTACGGCTTC TAGGCTAGAC ATTCTTAACG AAAACATCAA GGCGCGTAAT
 901 TTCACCTTCG AGCAAACCAA AGATAAAGCG CTCGCTGAAA TTGTGAATCA
 951 CGGTAAATT ACTGTGGTAA AAGACGGCAG TGTAAATCTT ATTGGTGGCA
 1001 AAGTGAAAA CGAGGGTGTG ATTAGCGTAA ATGGTGGCAG CATTCTTTA
 1051 CTCGCAGGGC AAAAATCAC CATCAGCGAT ATAATAAACCA ACCATTAC 11/68
 1101 TTACAGCATTT GCGGGCCTG AAAATGAAGC GGTCAATCTG GGGATATT
 1151 TTGCCAAAGG CGGTAACATT AATGTCCGTG CTGCCACTAT TCGAAACCAA
 1201 GGTAAACTTT CTGCTGATTCT TGTAAAGCAA GATAAAAGCG GCAATATTGT
 1251 TCTTCCGCC AAAGAGGGTG AAGCGGAAT TGGGGGTGTA ATTTCGGCTC
 1301 AAAATCAGCA AGCTAAAGGC GGCAAGGCTGA TGATTACAGG CGATAAAAGTC
 1351 ACATTTAAAA CAGGTGCAGT TATCGACCTT TCAGGTAAAG AAGGGGAGA
 1401 AACTTACCTT GCGGGTGCAG AGCGGGCGA AGGTAAAAAC GGCATTCAAT
 1451 TAGCAAAGAA AACCTCTTTA GAAAAGGCT CAACCATCAA TGTATCAGGC
 1501 AAAGAAAAAG GCGGACGGCGC TATTGTGTGG GGGGATATTG CGTTAATTGA

FIG. 3C.

1551 CGGCAATATT AACGGCTCAAG GTAGTGGTGA TATCGCTAAA ACCGGTGGTT
 1601 TTGTGGAGAC ATCGGGCAT TATTATCCA TTGACAGCAA TGCAATTGTT
 1651 AAAACAAAAG AGTGGTTGCT AGACCCTGAT GATGTAACAA TTGAAGCCGA
 1701 AGACCCCTT CGCAATAATA CCGGTATAAA TGATGATT CCAACAGGCA
 1751 CCGGTGAAGC AAGCGACCCCT AAAAAAATA GCGAACCTCAA AACAACGCTA
 1801 ACCAATAACAA CTATTCAAATTATCTGAAA AACGGCTGGT CAATGAATAAT
 1851 AACGGCATCA AGAAAACCTTA CGGTTAATAG CTCAAATCAAC ATCGGAAGCA
 1901 ACTCCCACCTT AATTCTCCAT AGTAAAGGTC AGCGTGGCGG AGGGGTTTCAG
 1951 ATTGATGGAG ATATTACTTC TAAAGGGGA AATTAAACCA TTATTTCTGG
 2001 CGGATGGTT GATGTTCATTA AAAATATTAC GCTTGATCAG GTTTTTTTAA
 2051 ATATTACCGC CGCTTCCGTA GCTTTTGAAG GTGAAATAA CAAAGCACGC
 2101 GACGGGGCAA ATGCTAAAT TGTGCCCCAG GGCACGTAA CCATTACAGG
 2151 AGAGGGAAAA GATTCAAGG CTAACAAACGT ATCTTTAAC GGAAACGGTA
 2201 AAGGTCTGAA TATCATTCA TCAGTGAATA ATTTAACCCA CAATCTTAGT
 2251 GGGCACAAATTAA ACATATCTGG GAATATAACA ATTAACCAAA CTACGAGAAA
 2301 GAACACCTCG TATTGGCAA CCAGCCATGA TTGGCAGTGG AACGTCAGTC
 2351 CTCTTAATCT AGAGACAGGCC GCAAATTAA CCTTTATTAA ATACATTCA

12 / 68

FIG. 3D.

2401 AGCAATAGCA AAGGCTTAAC AACACAGTAT AGAAGCTCTG CAGGGGTGAA
 2451 TTTTAACGGC GTAAATGGCA ACATGTCACTT CAATCTCAA GAAGGAGCGA
 2501 AAGTTAATT CAATTAAA CCAAAACGAGA ACATGAACAC AAGCAAACCT
 2551 TTACCAATTC GGTTTTAGC CAATATCACA GCCACTGGTG GGGCTCTGT
 2601 TTTTTTGAT ATATATGCC ACCATTCTGG CAGAGGGCT GAGTTAAAAA
 2651 TGAGTGAAAT TAATATCTCT AACGGCGCTA ATTTCACCTT AAATTCCCAT
 2701 GTTGGCGCG ATGACGGCTT TAAATCAAC AAAGACTTAA CCATAATGC 13 /
 2751 AACCAAATTCA AATTTCAGCC TCAGACAGAC GAAAGATGAT TTTTATGACG 68
 2801 GGTACGCCACG CAATGCCATC AATTCAAACCT ACAACATATC CATTCTGGC
 2851 GGTATGTCA CCCTTGGTGG ACAAAACTCA AGCAGCAGCA TTACGGGAA
 2901 TATTACTATC GAGAAAGCAG CAAATGTTAC GCTAGAAGCC AATAACGCC
 2951 CTAATCAGCA AAACATAAGG GATAGAGTTA TAAAACCTGG CAGCTTGCTC
 3001 GTTAATGGGA GTTTAAGTTT AACTGGCGAA AATGCAGATA TTAAAGGCAA
 3051 TCTCACTATT TCAGAAAGCG CCACTTTAA AGGAAAGACT AGAGATACCC
 3101 TAAATATCAC CGGCAATTTC ACCAATAATG GCACTGCCGA ATTAAATATA
 3151 ACACAAGGAG TGGTAAACT TGGCAATGTT ACCAATGATG GTGATTAAA

FIG. 3E.

3201 CATTACCACT CACGGCTAAC GCAACCAAAC AAGCATTAC GCGGGAGATA
 3251 TAATCAACAA AAAAGGAAGC TTAATATTAA CAGACAGTAA TAATGATGCT
 3301 GAAATCCAAA TTGGCGGCAA TATCTCGCAA AAAGAAGGCA ACCTCACGAT
 3351 TTCTTCCGAT AAAATTATA TCACCAAACA GATAACATC AAAAGGGTA
 3401 TTGATGGAGA GGACTCTAGT TCAGATGGGA CAGTAATGC CAACCTAACT
 3451 ATTAAACCA AAGAATTGAA ATTGACAGAA GACCTAAGTA TTTCAGGTT
 3501 CAATAAGCA GAGATTACAG CCAAAAGATGG TAGGAGATTAA CTATTGGCA
 3551 ACAGTAATCA CGGTAACAGC GGTGCCGAAG CCAAAACAGT AACTTTAAC
 3601 AATGTTAAAG ATTCAAAAT CTCTGCTGAC GGTCAACATG TGACACTAAA
 3651 TAGCAAAGTG AAAACATCTA GCAGCAATGG CGGACGTTGAA AGCAATAGCG
 3701 ACAACGATAC CGGCTTAAC ATTACTGCAA AAAATGTAGA AGTAACAAA
 3751 GATATTACTT CTCTCAAAAC AGTAAATATC ACCGGCTCGG AAAAGGTTAC
 3801 CACCACAGCA GGCTCGACCA TTAACGGCAAC AAATGGCAA GCAAGTATTA
 3851 CAACCAAAAC AGGTGATATC AGCGGTACGA TTTCCGGTAA CACGGTAAGT
 3901 GTTAGCGGA CTGGTGATT AACCACTAAA TCCGGCTCAA AAATTGAAGC
 3951 GAAATCGGGT GAGGCTAATG TAACAAAGTGC AACAGGTACA ATTGGGGTA

FIG. 3F.

4001 CAATTCCGG TAATACGGTA AATGTTACGG CAAACGGCTGG CGATTAAACA
 4051 GTTGGGAATG GCGCAGAAAT TAATGCGACA GAAGGGAGCTG CAACCTTAAC
 4101 CGCAACAGGG AATAACCTTGA CTACTGAAGC CGGTTCTAGC ATCACTTCAA
 4151 CTAAGGGTCA GGTAGACCTC TTGGCTCAGA ATGGTAGGCAT CGCAGGAAGC
 4201 ATTAATGCTG CTAATGTGAC ATTAATACT ACAGGGCACCT TAACCACCGT
 4251 GGCAGGGCTCG GATATTAAG CAACCAGGG CACCTGGGT ATTAAACGCCA
 4301 AAGATGCTAA GCTAAATGGT GATGCCATCAG GTGATAGTAC AGAAGTGAAT
 4351 GCAGTCAACG CAAGGGCTC TGGTAGTGTG ACTGGCGCAA CCTCAAGCAG
 4401 TGTGAATATC ACTGGGGATT TAAACACACT AAATGGGTT AATATCATT
 4451 CGAAAGATGG TAGAAACACT GTGCGCTTAA GAGGCAAGGA AATTGAGGTG
 4501 AAATATATCC AGCCAGGTGT AGCAAGTGTAA GAAGAAAGTAA TTGAAGCGAA
 4551 ACGGCGTCCTT GAAAAGTAA AAGATTATC TGATGAGGAA AGAGAAACAT
 4601 TAGCTAAACT TGGTGTAAAGT GCTGTACCGTT TTGTTGAGGCC AAATAATAACA
 4651 ATTACAGTCA ATACACAAA TGAATTACA ACCAGACCGT CAAGTCAAGT
 4701 GATAATTCTT GAAGGTAAGG CGTGTTCCTC AAGTGGTAAT GGGGCACCGAG
 4751 TATGTACCAA TGTTGCTGAC GATGGACAGC CGTAGTCAGT ATTGACAAG
 4801 GTAGATTTCATC TCCCTGCAATG AAGTCATTTT ATTTCCTGTT TATTACTGT

16/68

FIG. 3G.

4851	GTGGGTTAAA	GTTCAGTACCG	GGCTTTACCC	ATCTTGTAAA	AAATTACGGA
4901	GAATAACAATA	AAGTATTTTT	AACAGGGTAT	TATTATG	

FIG. 4A. AMINO ACID SEQUENCE OF HIGH MOLECULAR WEIGHT PROTEIN 2

1 MNK IYRIKFS KRLNALVAVS ELARGCDHST EKGSEKPARM KVRHLALKPL
 51 SAMILLSLGVT SIPQSVLASC LQGMDVVHGT ATMQVDGNNKT IIRNSVDAIT
 101 NWKQFNIDQN EMVQFLQENN NSAVFNRVTS NQISQLKGIL DSNGQVFLIN
 151 PNGITIGKDA IINTNGFTAS TLDISNENIK ARNFTFEQTK DKALAEIVNH
 201 GLITVGKDGS VNLLIGGKVKN EGVISVNGGS ISLLAGQKIT ISDIINPTIT
 251 YSIAAPENEAVNLDIFAKG GNINVRRAATI RNQGKLSADS VSKDKSGNIV
 301 LSAKEGEAEI GGVISAQNQQ AKGGKLMITG DVVTIJKTGAV IDLSGKEGGGE
 351 TYLGGDERGE GKNGIQIAKK TSLEKGSTIN VSGKEKGRA IVWGDIALID
 401 GNINAQGSGD IAKTGGFVET SGHDLFIKDN AIVDAKEWLL DFDNVSINAE
 451 DPLRNNTGIN DEFPTGTGEA SDPKKKNSELK TTLTNTTISN YLKNAWTMNI
 501 TASRKLTVNS SINIGSNSHL ILHSKGQRGG GVQIDGDTIS KGGNLTIYSG
 551 GWVDVHKNT LDQGFLNITA ASVAFEGGNN KARDAANAKI VAQGTVTITG
 601 EGKDFRANNV SLNGTGGKGLN IISSVNNLTH NLSGTINISG NITINQTTTRK
 651 NTSYWQTSHD SHWNVSAALNL ETGANFTFIK YISSNSKGGLT TQYRSSAGVN
 701 FNGVNGNMSF NLKEGAKVNF KLKPNNENMNT SKPLPIRFLA NITATGGGSV

17 / 60

FIG. 4B.

751 FFDIYANHSG RGAELKMSEI NISNGANFTL NSHVRGDDAF KINKDLTINA
 801 TNSNFSLRQT KDDFYDGYAR NAINSTYNIS ILGGNVTLGG QNSSSSITGN
 851 ITIEKAANVT LEANNAPNQQ NIRDRVIKLG SLLVNGSLSL TGENADIKGN
 901 LTISESATFK GKTRDTLNIT GNFTNNNGTAE INITQGVVKL GNVTNDGDLN
 951 ITTHAKRNQR SIIGGDIINK KGSLNITDSN NDAEIQIGGN ISQKEGNLTI
 1001 SSDKINITKQ ITIKKGIDGE DSSSDATSNA NLTIKTKELK LTEDLSISGF
 1051 NKAETAKDG RDLTIGNSND GNSGAEAKTV TFNNVKDSKI SADGHNVTLN
 1101 SKVKTSSNG GRESNSDNDT GLTITAKNVE VNKDITSLKT VNITASEKVT
 1151 TTAGSTINAT NGKASITTKT GDISGTTISGN TVSVSATVDL TTKSGSKIEA
 1201 KSGEANVTSA TGTIGGTISG NTVNVTANAG DLTVGNGAEI NATEGAATLT
 1251 ATGNTLTEA GSSITSTKGQ VDLAQNGSI AGSINAANVT LNTTGTLLTV
 1301 AGSDIKATSG TLVINAKDAK LNGDASGDEST EVNAVNASGS GSVTAATSSS
 1351 VNITGDLNTV NGLNIIISKDG RNTVRLRGKE LEVKYIOPGV ASVVEEVIEAK
 1401 RVLEVKVDLS DEERETLAKL GVSAYRFVEP NNTITVNTQN EFTTRPSSQV
 1451 TISEGKACFS SGNGARVCTN VADDGQP

19/68

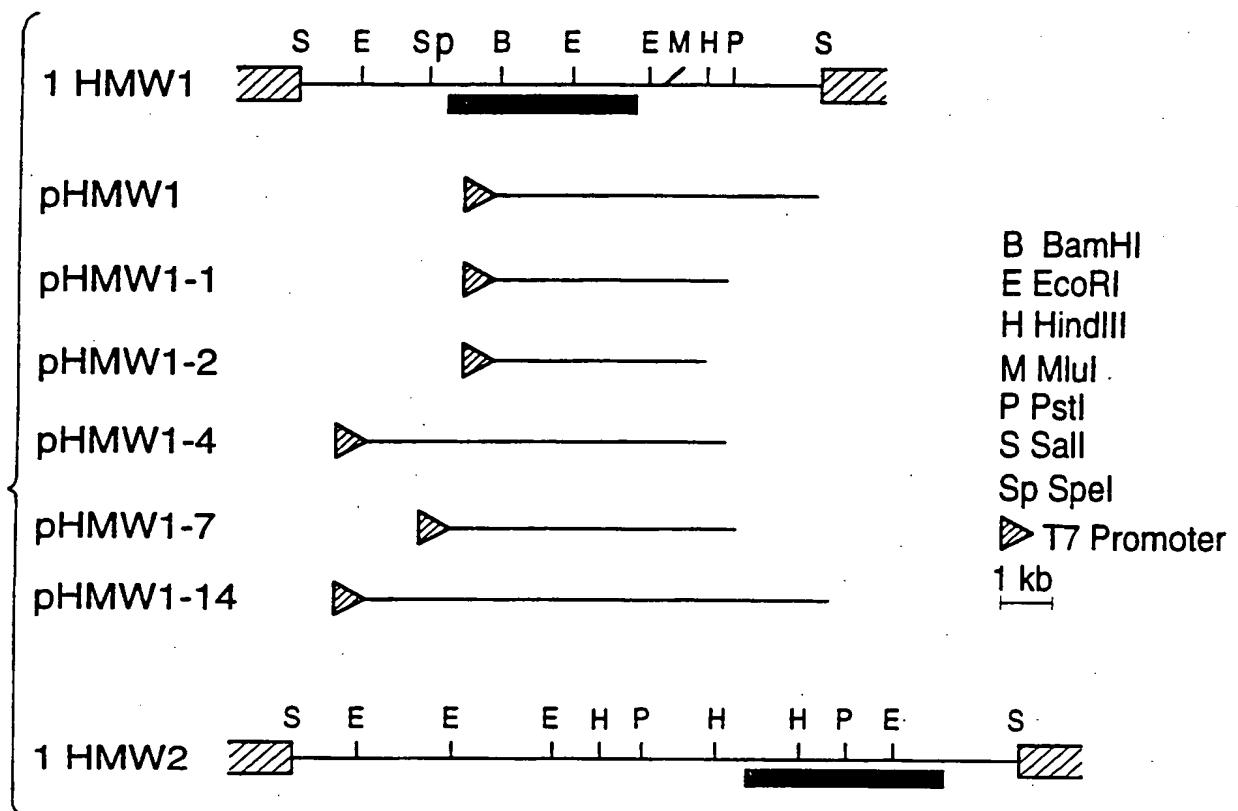


FIG. 5 A.

20/68

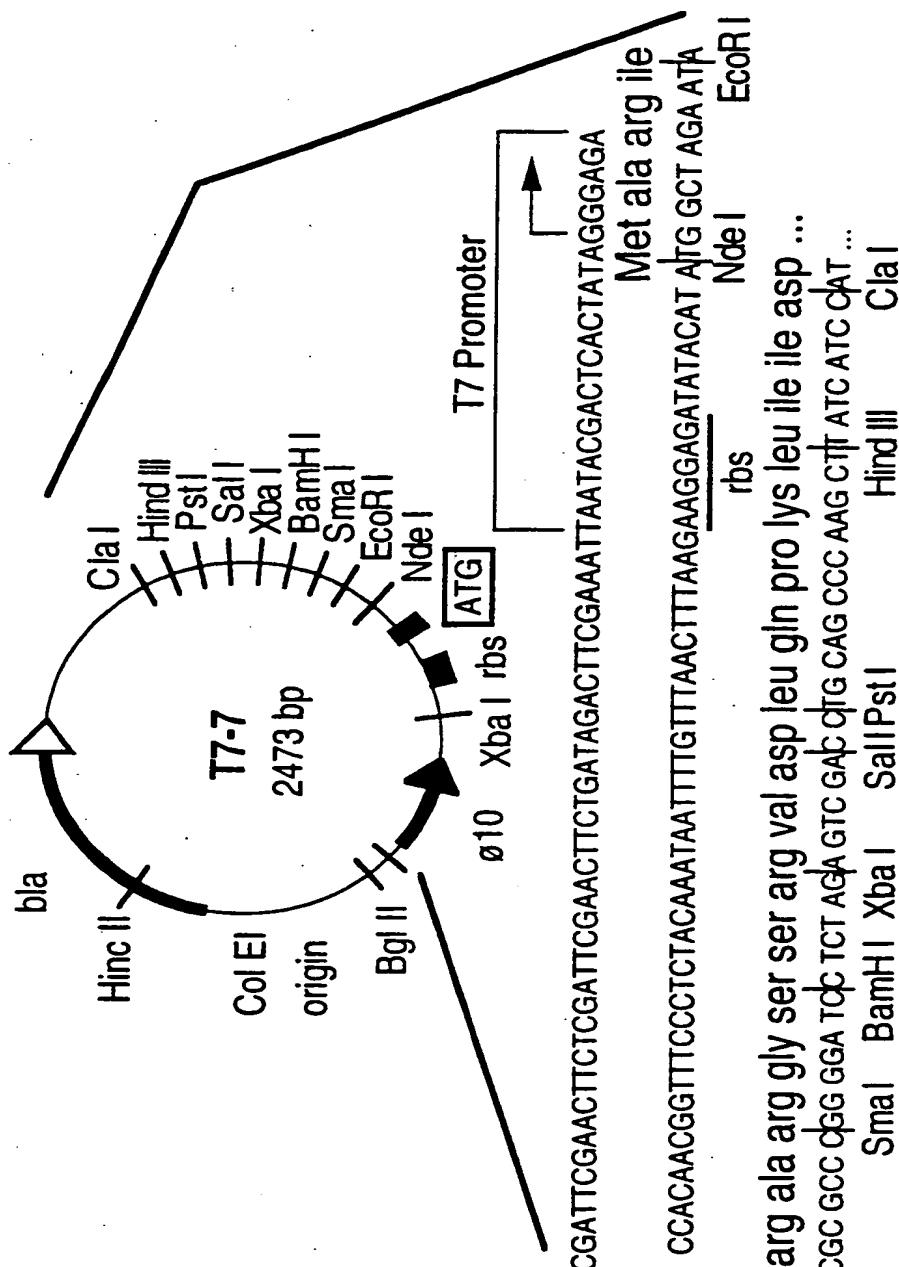


FIG. 5 B.

(A) Partial restriction maps of representative HMW1 and HMW2 recombinant phage and of HMW1 plasmid subclones. The shaded boxes indicate the locations of the structural genes. In the recombinant phage, transcription proceeds from left to right for the HMW1 gene and from right to left for the HMW2 gene. The methods used for construction of the plasmids shown are described in the text. (B) Restriction map of the T7 expression vector pT7-7. This vector contains the T7 RNA polymerase promoter ϕ 10, a ribosome - binding site (rbs), and the translational start site for the T7 gene 10 protein upstream from a multiple cloning site (37).

FIG. 6A.

1 ACAGCGTTCT CTTAATACTA GTACAAACCC ACAATAAAAT ATGACAAACA
 51 ACATTACAA CACCTTTTT GCAGTCTATA TGCAAATATT TTAAAAAATA
 101 GTATAATCC GCCATATAAA ATGGTATAAT CTTTCATCTT TCATCTTCA
 151 TCTTCATCTT TCATCTTTTC ATCTTTCATC TTTCATCTT CATCTTCA
 201 CTTTCATCTT TCATCTTCA TCTTTCATCTT TTCATCTTC ACATGAAATG
 251 ATGACCGAG GGAAGGGAGG GAGGGCAAG AATGAAGAGG GAGGCTGAACG
 301 AACGCAAATG ATAAAGTAAT TAAATTGTTTC AACTAACCTT AGGAGAAAT /
 351 ATGACAAAGA TATATCGTCT CAAATTCAGC AAACGCCCTGA ATGCTTTGGT
 401 TGCTGTGTCT GAATTGGCAC GGGTTGTGA CCATTCCACA GAAAAGGCA
 451 GCGAAAAACC TGCTCGCATG AAAGTGGCTC ACTTAGCGTT AAAGCCACTT
 501 TCCGCTATGT TACTATCTT AGGTGTAACA TCTATTCCAC AATCTGTTT
 551 AGCAAGGGC TTACAAGGAA TGGATGTAGT ACACGGCACA GCCACTATGC
 601 AAGTAGATGG TAATAAACC ATTATCCGCA ACAGTGTGTA CGCTATCATT
 651 AATTGGAAC AATTAAACAT CGACCAAAAT GAAATGGTGC AGTTTTTACA
 701 AGAAAACAAAC AACTCCGCCG TATTCAACCG TGTACATCT AACCAAATCT
 751 CCCAATTAAA AGGGATTITA GATTCTAACG GACAAGTCTT TTTAATCAAC

FIG. 6B.

801 CCAAATGGTA TCACAATAGG TAAAGACGCA ATTATTAAACA CTAATGGCTT
 851 TACGGCTCTT ACGCTAGACA TTTCATAACGA AACATCAAG GCGCGTAATT
 901 TCACCTCGA GCAAACAAA GATAAAGCGC TCGCTGAAAT TGTGAATCAC
 951 GGTTTAATTA CTGTCGGTAA AGACGGCAGT GTAAATCTTA TTGCTGGCAA
 1001 AGTGAACAC GAGGGTGTGA TTAGCGTAAA TGCTGGCAGC ATTFTCTTTAC
 1051 TCGCAGGGCA AAAATCACC ATCAGCGATA TAATAAACCC AACCATTA
 1101 TACAGCATTG CCGGCCCTGA AATGAAGCG GTCAATCTGG GCGATATT'TT
 1151 TGCCAAGGC GGTAAACATTA ATGTCGGTGC TGCCACTATT CGAACCAAG
 1251 CTTCCGCCA AAGAGGGTGA AGCGGAATT GGGGGTAA TTTCCGCTCA
 1301 AAATCAGCAA GCTAAAGGGC GCAAGGCTGAT GATTACAGGC GATAAAGTCA
 1351 CATTAAAC AGGTGCAGTT ATCGACCTTT CAGGTAAAGA AGGGGGAGAA
 1401 ACTTACCTTG GCGGTGACGA GCGCGGGCAA GTAAAAACG GCATTCAAATT
 1451 AGCAAAAGAAA ACCTCTTTAG AAAAAGGCTC AACCATCAAT GTATCAGGCA
 1501 AAGAAAAGG CGGACGCGCT ATTGTGTGGG GCGATATTGC GTTAATTGAC
 1551 GGCAATATTA ACGCTCAAGG TAGTGGTGAT ATCGCTAAA CCGGTGGTTT
 1601 TGTGGAGACG TCGGGGCATG ATTATTCAT CAAAGACAAAT GCAAGACAAAT

22/68

FIG. 6C.

1651 ACGCCAAAGA GTGGTTGTTA GACCCGGATA ATGTATCTAT TAATGCAGAA
 1701 ACAGCAGGAC GCAGCAATAC TTCAGAAGAC GATGAATAACA CGGGATCCGG
 1751 GAATAGTGCC AGCACCCAA AACGAAACAA AGAAAAGACA ACATTAACAA
 1801 ACACAACTCT TGAGAGTATA CTAAAAAAAG GTACCTTTGT TAACATCACT
 1851 GCTAATCAAC GCATCTATGT CAATAGCTCC ATTAAATTAT CCAATGGCAG
 1901 CTTAACTCTT TGGAGTGAGG GTCGGAGCGG TGGGGCGGT GAGATTAACA
 1951 ACGATATTAC CACCGGTGAT GATACCAGAG GTGCCAAACTT AACAAATTAC 23 / 68
 2001 TCAGGGGCT GGGGTGATGT TCATAAAAAT ATCTCACTCG GGGCCAAAGG
 2051 TAACATAAAC ATTACAGCTA AACAAAGATAAT CGCCTTGTGAG AAAGGAAGCA
 2101 ACCAAGTCAT TACAGGTCAA GGGACTATTAA CCTCAGGCCA TCAAAAGGT
 2151 TTTAGATTAA ATAATGTCTC TCTAAACGGC ACTGGCAGCG GACTGCAATT
 2201 CACCACTAAA AGAACCAATA AATACGCTAT CACAAATAAA TTTGAAGGGA
 2251 CTTAAATAT TTCAGGGAAA GTGAACATCT CAATGGTTT ACCTAAAAAT
 2301 GAAAGTGGAT ATGATAAAAT CAAAGGACGC ACTTACTGGA ATTTAACCTC
 2351 GAAAGTGGAT ATGATAAAAT CAAAGGACGC CCTCACTATT GACTCCAGAG
 2401 GAAGCCATAG TGCAAGGCACA CTTACCCAGC CTTATAATT AAACGGTATA
 2451 TCATTCAACA AAGACACTAC CTTAAATGTT GAACGAAATG CAAGAGTCAA

FIG. 6D.

2501 CTTTGACATC AAGGCACCAA TAGGGATAAA TAAGTATTCT AGTTTGAAATT
 2551 ACGCATCAT TAATGGAAAC ATTCAGTTC CGGGAGGGGG GAGTGTGTGAT
 2601 TTCACACTTC TCGCCTCATC CTCTAACGTC CAAACCCCCG GTGTAGTTAT
 2651 AAATTCTAAA TACTTTAATG TTTCAACAGG GTCAAGTTA AGATTAAAAA
 2701 CTTCAGGCTC AACAAAAACT GGCTTCTCAA TAGAGAAAGA TTTAACTTTA
 2751 AATGCCACCG GAGGCCAACAT AACACTTTTG CAAGTGTGAG GCACCGATGG
 2801 AATGATTGGT AAGGCATTG TAGCCAAAAA AACACATAACC TTTGAAGGAG 24/68
 2851 GTAAAGATGAG GTTGGCTCC AGGAAAGCCG TAACAGAAAT CGAAGGCAAT
 2901 GTTACTATCA ATAACAACGC TAACGTCACT CTTATCGGTT CGGATTGTGA
 2951 CAACCATCAA AACCTTTAA CTATTAAGAA AGATGTCATC ATTAATAGCG
 3001 GCAACCTTAC CGCTGGAGGC AATATTGTCA ATATAGCCGG AAATCTTACC
 3051 GTTGAAGTA ACGCTAATT CAAAGCTATC ACAAAATTCA CTTTTAATGT
 3101 AGGGGGCTTG TTGACAAACA AGGCAATTCA AAATATTCC ATTGCCAAG
 3151 GAGGGGCTCG CTTAAAGAC ATTGATAATT CCAAGAATT AAGCATCACC
 3201 ACCAACTCCA GCTCCACTTA CCGCACTATT ATAAGCGGCA ATATAACCAA
 3251 TAAAAACGGT GATTTAAATA TTACGAACGA AGGTAGTGTACTGAAATGC

FIG. 6E.

3301 AAATGGCGG CGATGTCTCG CAAAAGAAG GTATCTCAC GATTTCTTCT
 3351 GACAAATCA ATATTACCA ACAGATAACA ATCAAGGCAG GTGTTGATGG
 3401 GGAGAATTCC GATTCAAGCG CGACAAACAA TGCCAATCTA ACCATTA
 3451 CCAAGAATT GAAATTAACG CAAGACCTAA ATATTCAGG TTTCAATA
 3501 GCAGAGATTA CAGCTAAAGA TGGTAGTGAT TTAACTATTG GAAACACCAA
 3551 TAGTGCTGAT GGTACTAATG CCAAAAAGT AACCTTTAAC CAGGTTAAAG
 3601 ATTCAAAAT CTCTGCTGAC GGTACACAAGG TGACACTACA CAGCAAAGTG
 3651 GAAACATCCG GTAGTAATAA CAACACTGAA GATAGCAGTG ACAATAATGC
 3701 CGGCTTAAC TATCGATGCAA AAAATGTAAC AGTAAACAA ATATTA
 3751 CTCACAAAGC AGTGAGGATC TCTGGCACAA GTGGAGAAAT TACCACTAAA
 3801 ACAGGTACAA CCATTAACGC AACCACTGGT AACGTTGGAGA TAACCGCTCA
 3851 AACAGGTAGT ATCCTAGGTG GAATTGAGTC CAGCTCTGGC TCTGTAACAC
 3901 TTACTGCCAAC CGAGGGCGCT CTTGCTGTAA GCAATATTTC GGGCAACACC
 3951 GTTACTGTAA CTGCAAATAG CGGTGCATTA ACCACTTGG CAGGCTCTAC
 4001 ATTAAAGGA ACCGAGAGTG TAACCACTTC AAGTCAATCA GGCAGATATCG
 4051 GCGGTACGAT TTCTGGTGGC ACAGTAGAGG TTAAAGCAAC CGAAAGTTA

FIG. 6F.

4101 ACCACTCAAT CCAATTCAA AATTAAGCA ACAACAGGG AGGCTAACGT
 4151 AACAAAGTGC A ACAGGTACAA TTGGTGGTAC GATTTCGGT AATACGGTAA
 4201 ATGTTACGGC AACAGCTGGC GATTTAACAG TTGGAATGG CGCAGAAATT
 4251 AATGCGACAG AAGGAGCTGC AACCTTAAC ACATCATCGG GCAAATTAAC
 4301 TACCGAAGCT AGTTCACACA TTACTTCAGC CAAGGGTCAG GTAAATCTTT
 4351 CAGCTCAGGA TGGTAGCGTT GCAGGAAGTA TTATGCCGC CAATGTGACA
 4401 CTAAAACTA CAGGCACTT AACTACCGTG AAGGGTCAA ACATTAATGC
 4451 AACCAGGGT ACCCTGGTTA TTAACGCCAA AGACGCTGAG C²⁶ TAAATGGCG
 4501 CAGCATTGGG TAAACCACACA GTGGTAAATG CAACCAACGC AAATGGCTCC
 4551 GGCAGCGTAA TCCGACAAAC CTCAAGCAGA GTGAACATCA C⁶⁸ TGGGGATT
 4601 AATCACAAATA AATGGATTAA ATATCATTTCA AAAAACGGT ATAAACACCG
 4651 TACTGTTAAA AGGCAGTTAAA ATTGATGTGA AATACATTCA ACCGGGTATA
 4701 GCAAGCGTAG ATGAAGTAAT TGAAGCGAAA CGCATCCTTG AGAAGGTAAA
 4751 AGATTATCT GATGAAGAAA GAGAAGCGTT AGCTAAACTT GCGGTAAGTG
 4801 CTGTCAGTT TATTGAGCCA ATAATACAA TTACAGTCGA TACACAAAT
 4851 GAATTGCAA CCAGACCATT AAGTCCAATA GTGATTCTG AGGCAGGGC
 4901 GTGTTCTCA AACAGTGATG GCGCGACGGT GTGCGTTAAAT ATCGCTGATA

FIG. 6G.

4951 ACGGGCGGT A GCGGTCAGTA ATTGACAAGG TAGATTTCAT CCTGCAAATGA
 5001 AGTCATTTA TTTTCGTATT ATTTACTGTG TGGTTAAAG TTCAGTACGG
 5051 GCTTACCCA TCCTGTAAA ATTACGGAG AATACAATAA AGTATTTTA
 5101 ACAGGTTATT ATTATGAAA ATATAAAAG CAGATTAAA CTCAGTGCAA
 5151 TATCACTATT GCTTGGCCTG GCTTCTTCAT CATTGTATGC AGAAGAAGCG
 5201 TTTTAGTAA AAGGCTTCA GTTATCTGGT GCACTTGAAA CTTTAAGTGA
 5251 AGACGCCAA CTGCTCTGTAG CAAATCTT ATCTAAATAC CAAGGCTCGC 27/68
 5301 AACCTTAAC AAAACCTAAA ACAGCACAGC TTGAATTACA GGCTGTGCTA
 5351 GATAAGATIG AGCCAATAA GTTGTATGTG ATATTGCCAC ACAAAACCAT
 5401 TACGGATGGC AATATTATGT TTGAGCTAGT CTCGAAATCA GCCGCAGAAA
 5451 GCCAAGTTT TTATAAGGCG AGCCAGGGTT ATAGTGAGA AAATATCGCT
 5501 CGTAGCCTGC CATCTTGAA ACAAGGAAA GTGTATGAAG ATGGTCGTCA
 5551 GTGGTTTCGAT TTGCGTGAAT TCAATATGGC AAAAGAAAAT CCACTTAAAG
 5601 TCAC'TCGGGT GCATTACGAG TTAAACCCCTA AAAACAAAC CTCCTGATTG
 5651 GTAGTTGCAG GTTTTCGCC TTTTGGCAAAC ACGCGTAGCT TTGTTCCCTA
 5701 TGATAAATTTC GCGCAAGGG AGTTAACTA TCAACGTGTA AGTCTAGGTT

FIG. 6H.

5751 TTGTAAATGC CAATTGACCC GGACATGATG ATGTATTAAA TCTAACGCCA
 5801 TTGACCAATG TAAAGCACC ATCAAATCT TATGCGGTAG GCATAGGATA
 5851 TACTTATCCG TTTATGATA AACACCAATC CTTAAGTCTT TATACCAGCA
 5901 TGAGTTATGC TGATTCTAAT GATATCGACG GCTTACCAAG TCGGATTAAAT
 5951 CGTAAATTAT CAAAGGTCA ATCTATCTCT GCGAATCTGA AATGGAGTTA
 6001 TTATCTCCCG ACATTTAACCC TTGGAATGGA AGACCAGTTT AAAATTAATT
 6051 TAGGCTACAA CTACCGCCAT ATTAATCAA CATCCGAGTT AACACCCCTG
 6101 GGTGCAACGA AGAAAAATT TGCAGTATCA GCGGTAAGTG CAGGCATTGA 28/68
 6151 TGGACATATC CAATTACCC CTAACAAAT CTTTAATATT GATTAACTC
 6201 ATCATTATTA CGCGAGTAA TTACCGGGCT CTTTTGGAAT GGAGGCCATT
 6251 GCGGAAACAT TTAATCGCAG CTATCACATT AGCACAGCCA GTTTAGGGTT
 6301 GAGTCAGAG TTTGCTCAAG GTTGGCATTT TAGCAGTCAA TTATCGGGTC
 6351 AGTTTACTCT ACAAGATATA AGTAGGCATAG ATTATTCCTC TGTAAACAGGT
 6401 ACTTATGGCG TCAGAGGCCT TAAATACGGC GGTGCAAGTG GTGAGGGCGG
 6451 TCTTGTATGG CGTAAATGAAT TAAGTATGCC AAAATAACACC CGCTTTCAAA
 6501 TCAGCCCTTA TGGTTTTAT GATGCAGGTC AGTTCCGGTTA TAATAGCGAA
 6551 AATGCTAAA CTTACGGCGA AGATATGCAC ACGGTATCCCT CTGCGGGTT

FIG. 6I.

6601	AGGCATTAAC	ACCTCTCTA	CACAAACTT	AAGCTTAGAT	GCTTTGTTG	
6651	CTCGTGGCTT	TGCCAATGCC	AATAGTGACA	ATTGAAATGG	CAACAAAAAA	
6701	CGCACAAAGCT	CACCTACAAAC	CTTCTGGGT	AGATTAAACAT	TCAGTTCTTA	
6751	ACCCTGAAAT	TTAATCAACT	GGTAAGCGTT	CCGCCTACCA	GTTTATAACT	
6801	ATATGCTTTA	CCGCCATT	TACAGTCTAT	ACGCAACCT	GTGTTCATCC	
6851	TTATATATCA	AACAAACTAA	GCAAACCAAG	CAAACCAAGC	AAACCAAGCA	
6901	AACCAAGCAA	ACCAAGCAA	CCAAGCAAAC	CAAGCAAACC	AAGCAAACCA	29
6951	AGCAAACCAA	GCAAACCAAG	CAAACCAAGC	AAACCAAGCA	ATGCTAAAAA	68
7001	ACAATTATA	TGATAAACTA	AAACATATACTC	CATACCATGG	CAATACAAGG	
7051	GATTAAATA	TATGACAAAA	GAAAATTAC	AAAGTGTTC	ACAAAATAACG	
7101	ACCGCTTCAC	TTGTAGAATC	AAACAAACGAC	CAAACTTCCC	TGCAAATACT	
7151	TAAACAAACCA	CCCAAACCCA	ACCTATTACG	CCTGGAACAA	CATGTGGCCA	
7201	AAAAGATTA	TGAGCTTGCT	TGCCGCGAAT	TAATGGCGAT	TTTGGAAAAA	
7251	ATGGACGGCTA	ATTGGAGG	CGTTCAACGAT	ATTGAATTTG	ACGCACCTG	
7301	TCAGCTGGCA	TATCTACCCG	AAAAACTACT	AATTCAATT	GCCACTCGTC	
7351	TCGCTTAATGC	AATTACAAACA	CTCTTTCCG	ACCCCGAATT	GGCAATTTC	

FIG. 6J.

7401 GAAGAAGGG CATTAAAGAT GATTAGCCTG CAA CGCTGGT TGACGCTGAT
 7451 TTTGCCTCT TCCCCCTACG TTAACGGAGA CCATATTCTC AATAAAATA
 7501 ATATCAACCC AGATTCCGAA GGTGGCTTTC ATTAGAAC AGACAACTCT
 7551 TCTATTGCTA AATTCTGTAT TTTTTACTTA CCCGAATCCA ATGTCATAAT
 7601 GAGTTTAGAT GCGTTATGGG CAGGGAAATCA ACAACTTTGT GCTTCATTTGT
 7651 GTTTGCGTT GCAGTCTCA CGTTTTATTC GTACTGCATC TGC GTTTCAT
 7701 AAAAGAGCGG TGGTTTACA GTGGTTTCCT AAAAAACTCG CCGAAATTGC 30
 7751 TAATTAGAT GAATTGCCTG CAAATATCCT TCATGATGTA TATATGCACT 68
 7801 GCAGTTATGA TTAGCAAAA ACAAGGCACG ATGTTAACG TCCATTAAAC
 7851 GAACTTGTCC GCAAGCATA CCTCACCGCAA GGATGGCAAG ACCGCTACCT
 7901 TTACACCTTA GGTAAAAGG ACGGCAAACC TGTGATGATG GTACTGCTTG
 7951 AACATTTAA TTCGGGACAT TCGATTATC GCACGCATC AACTTCAATG
 8001 ATTGCTGCTC GAGAAAATT CTATTAGTC GGCTTAGGCC ATGAGGGCGT
 8051 TGATAACATA GGTCGAGAAG TGTGGACGA GTTCTTTGAA ATCAGTAGCA
 8101 ATAATATAAT GGAGAGACTG TTTCATCC GTAAACAGTG CGAAACTTTC
 8151 CAACCCGGCAG TGTCTTATC GCCAAGGCATT GGCATGGATA TACACGAT

FIG. 6K.

8201 TTTTGTGAGC AACACTCGGC TTGCCCCCTAT TCAAGCTGTA GCCTTGGGTCA
 8251 ATCCTGCCAC TACGCATTCT GATTATTATTG ATTATGTCAT CGTAGAACAT
 8301 GATTATGTGG GCAGTGAAGA TTGTTAGC GAAACCCTTT TACGCTTACCA
 8351 CAAAGATGCC CTACCTTATG TACCATCTGC ACTCGCCCCA CAAAAGTGG
 8401 ATTATGTACT CAGGGAAAC CCTGAAGTAG TCAATATCGG TATTGCCGCT
 8451 ACCACAAATGA ATTAAACCC TGAATTTTTG CTAACATTGC AAGAAATCAG
 8501 AGATAAAGCT AAAGTCAAAA TACATTTCAT TTTCGCACTT GGACAAATCAA
 8551 CAGGCTTGAC ACACCCATT GTCAAATGGT TTATCGAAAG CTATTAGGT^{31/68}
 8601 GACGGATGCCA CTGCACATCC CCACGCACCT TATCACGATT ATCTGGCAAT
 8651 ATTGGCGTGAT TGGGATATGC TACTAAATCC GTTCCCTTTC GGTAAATACTA
 8701 ACGGCATAAT TGATATGGTT ACATTAGGT TAGTTGGTGT ATGCAAAACCG
 8751 GGGGATGAAAG TACATGAACA TATTGATGAA GGTCTGTTTA AACGCTTAGG
 8801 ACTACCAGAA TGGCTGATAG CCGACACACG AGAAACATAT ATTGAATGTC
 8851 CTTTGCCTCT AGCAGAAAC CATCAAGAAC GCCTTGAACT CCGTCGTTAC
 8901 ATCATAGAAA ACAACGGCTT ACAAAAGCTT TTTACAGGGC ACCCTCGTCC
 8951 ATTGGGCCAA ATACTGCTTA AGAAAACAAA TGAATGGAAG CGGAAGCACT
 9001 TGAGTAAAAA ATAACGGTTT TTAAAGTAA AAGTGGGTAA AATTTCAAA

32 / 68

FIG. 6L.

9051 GCGTTTTAAA AACCTCTCAA AAATCAACCG CACTTTTATC TTTATAACGC
9101 TCCCGGGCGC TGACAGTTA TCTCTTTCTT AAAATAACCA TAAAATTGTG
9151 GCAATAGTTG GGTAAATCAA TTCAATTGTT GATAACGGCAA ACTAAAGACG
9201 GCGCGTTCTT CGGGCAGTCAT C

FIG. 7A.

1 CGCCCACTCA ATTTGGATT GTTGAATTCA AACTAACCAA AAAGTGGGT
 51 TAAATCTGT GGAGAAATA GGTTGTAGTG AAGAACGAGG TAATTGTTCA
 101 AAAGATAAA GCTCTCTAA TTGGCATATG GTTGGCGTTT CTTTTCGGT
 151 TAATAGTAA TTATATTCTG GACGACTATG CAATCCACCA ACAACTTTAC
 201 CGTTGGTTT AAGCGTTAAT GTAAAGTTCTT GCTCTTCTTG GCGAACATACGT
 251 AATCCCATTT TTTGTTAGC AAGAAAATGA TCGGGATAAT CATAATAGGT
 301 GTTGCACAA AATAAATTGATGTTCTAA AATCATAAAT TTTGCAAGAT
 351 ATTGTGGCAA TTCAAAACCT ATTGTGGCG AAATGCCAA TTTTAATTCA
 401 ATTCTTGTAA GCATAATATT TCCCACCTCAA ATCAAACCTGGT TAAATATA
 451 AGATAATAAA AATAAATCAA GATTTTGTG ATGACAAACA ACAATTACAA
 501 CACCTTTT GCAGTCTATA TGCAAAATATT TTAAAAAAAT AGTATAAATC
 551 CGCCATATAA AATGGTATAA TCTTCATCT TTCATCTTC ATCTTCATC
 601 TTTCATCTT CTCATCTT CTTTCATCTT TCATCTTC TCTTCATCT
 651 TTCACTCTTC ATCTTCATC TTTCATCTT CACATGAAAT GATGAACCGA
 701 GGGAAAGGAG GGAGGGCAA GAATGAAGAG GGAGCTGAAC GAAACGCAAAT
 751 GATAAACTAA TTAAATTGTT CAACTAACCT TAGGAGAAAA TATGAACAAAG

33 / 68

FIG. 7B.

801 ATATATCGTC TCAAATTCA G CAAACGCCCTG AATGCTTTGG TTGCTGTGTC
 851 TGAATTGGCA CGGGGTTGTG ACCATTCCAC AGAAAAAGGC AGCGAAAAAC
 901 CTGCTCGCAT GAAAGTGCCT CACTTAGCGT TAAAGCCACT TTCCGCTATG
 951 TTACTATCTT TAGGTGTAAC ATCTATTCCA CAATCTGT TT TAGCAAGCGG
 1001 CAATTAAACA TCGACCAAAA TGAATGGTG CAGTTTTAAC AAGAAAACAA
 1051 GTAAATAAAC CATTATCCGC AACAGTGTG ACGCTATCAT TAATTGGAAA
 1101 CAATTAAACA TCGACCAAAA TGAATGGTG CAGTTTTAAC AAGAAAACAA
 1151 CAACTCCGCC GTATTCAACC GTGTTACATC TAACCAAATC TCCCAATTAA 34 / 68
 1201 AAGGGATT AGATTCTAAC GGACAAGTCT TTTTAATCAA CCCAAATGGT
 1251 ATCACAAATAG GTAAAGACGC ATTATTAAC ACTAATGGCT TTACGGCTTC
 1301 TACGGCTAGAC ATTCTAACG AAAACATCAA GGGCGTAAAT TTCACCTTCG
 1351 AGCAAACCAA AGATAAAGCG CTCGCTGAAA TTGTTGAATCA CGGGTTAAATT
 1401 ACTGTGGTA AAGACCGGCAG TGTAAATCTT ATTGGTGGCA AAGTGAAGAA
 1451 CGAGGGTGTG ATTAGCGTAA ATGGTGGCAG CATTCTTTA CTCGCAGGGC
 1501 AAAAATCAC CATCAGCCGAT ATAATAAAC CAAACCATTAC TTACAGCATT
 1551 GCCGGGCCCTG AAAATGAAGC GGTCAATCTG GGCGATATT TTGCCAAAGG

FIG. 7C.

1601 CGGTAACATT AATGTCCTG CTGCCACTAT TCGAAACCAA GGTAAACTTT
 1651 CTGCTGATTCT TGTAAAGAAA GATAAAAGCC GCAATATTGT TCTTTCCGCC
 1701 AAAGAGGGTG AAGCGGAAAT TGGCGGTGTA ATTTCGGCTC AAAATCAGCA
 1751 AGCTAAAGGC GGCAGGTGA TGATTACAGG CGATAAAAGTC ACATTAAGAA
 1801 CAGGTGCAGT TATCGACCTT TCAGGTAAG AAGGGGGAGA AACTTACCTT
 1851 GCGGGTGCAG AGCGGGCGA AGGTAAAAAC GGCATTCAAT TAGCAAAGAA
 1901 AACCTCTTA GAAAAGGCT CAACCATCAA TGTATCAGGC AAAGAAAAG
 1951 GCGGACGGCG TATTGTGTGG GGGGATATTG CGTTAATTGA CGGCAATTATT³⁵
 2001 AACGGCTCAAG GTAGTGGTGA TATCGCTAAA ACCGGTGGTT TTGTGGAGAC⁶⁰
 2051 ATCGGGGCAT TATTATCCA TTGACAGCAA TGCAATTGTT AAAACAAAAG
 2101 AGTGGTGCT AGACCCTGAT GATGTAACAA TTGAAGCCGA AGACCCCTT
 2151 CGCAAATAATA CCGGTATAAA TGATGAATTCC CCAACAGGCA CCGGTGAAGC
 2201 AAGGGACCCCT AAAAAAATA GCGAACTCAA AACAAACGCTA ACCAATACAA
 2251 CTATTTCAAAT TATCTGAAA AACGCCCTGGA CAATGAATAT AACGGCATCA
 2301 AGAAAACCTTA CCGTTAATAG CTCAATCAAC ATCGGAAGCA ACTCCCACTT
 2351 ATTCTCCAT AGTAAAGGTC AGCGTGGGG AGGGCTTCAG ATTGATGGAG
 2401 ATATTACTTC TAAAGGGCA ATTAAACCA TTTATTCTGG CGGATGGGT

FIG. 7D.

2451 GATGTTCAT AAAATATTAC CCTTGATCAG GGTTTTTAA ATATTACCGC
 2501 CGCTTCCGTA CCTTTTGAAG GTGGAAATAA CAAAGCACGC GACGGGGCAA
 2551 ATGCTAAAT TGTGCCAG GGCAC'TGTAA CCATTACAGG AGAGGGAAA
 2601 GATTTCAGGG CTAACAAACGT ATCTTTAAC GGAAACGGTA AAGGTCTGAA
 2651 TATCATTCA TCAGTGAATA ATTTAACCCA CAATCTTAGT GGCACAAATT
 2701 ACATATCTGG GAATATAACA ATTAACCAA CTACGAGAAA GAACACCTCG
 2751 TATTGGCAA CCAGCCATGA TTTCGCACTGG AACGTCAGTG CTCCTTAATCT 36
 2801 AGAGACAGGC GCAAATTAA CCTTTTAA ATACATTICA AGCAATAGCA 68
 2851 AAGGCTTAAC AACACAGTAT AGAACGCTCTG CAGGGGTGAA TTTTAAACGGC
 2901 GTAAATGGCA ACATGTCATT CAATCTCAA GAAGGAGCGA AAGTTAATT
 2951 CAAATTAAA CCAAACGAGA ACATGAACAC AAGCAAACCT TTACCAATTIC
 3001 GGTTTTAGC CAATATCACA GCCAC'TGGTG GGGGCTCTGT TTTTTTGAT
 3051 ATATATGCCA ACCATTCTGG CAGAGGGCT GAGTTAAAAA TGAGTGAAT
 3101 TAATATCTT AACGGCGCTA ATT'TTACCTT AAATTCCCAT GTTGGGGCG
 3151 ATGACGCTTT TAAATCAAC AAAGACTTAA CCATAAATGC ACCAAATTCA
 3201 AATTTCAGGCC TCAGACAGAC GAAAGATGAT TTTATGACG GGTACGCACG

FIG. 7E.

3251 CAATGCCATC AATTCAACCT ACAACATATC CATTCTGGGC GTAAATGTCA
 3301 CCCTTGGTGG ACAAAACTCA AGCAGCAGCA TTACGGGGAA TATTACTATC
 3351 GAGAAAGCAG CAAATGTTAC GCTAGAAGCC AATAACGGCC CTAATCAGCA
 3401 AACACATAAAGG GATAGAGTTA TAAAACCTTGG CAGCTTGCTC GTAAATGGGA
 3451 GTTTAAGTTT AACTGGCGAA AATGCAGATA TAAAGGCAA TCTCACTATT
 3501 TCAGAAAGCG CCACTTTAA AGGAAAGACT AGAGATAACCC TAAATATCAC
 3551 CGGCAATTTC ACCAATAATG GCACGTGCCGA AATTAATATA ACACAGGAG
 3601 TGGTAAACT TGGCAATGTT ACCAATGATG GTGATTAAA CATTACCACT
 3651 CACGCTAAC GCAACCAAG AAGCATCATC GGGGAGATA TAATCAACAA
 3701 AAAAGGAAGC TAAATATTA CAGACAGTAA TAATGATGCT GAAATCCAA
 3751 TTGGGGCAA TATCTCGCAA AAAGAAGGCA ACCTCACCGAT TTCTTCCGAT
 3801 AAAATTAAATA TCACCAAACA GATAACAATC AAAAAGGGTA TTGATGGAGA
 3851 GGACCTCTAGT TCAGATGCGA CAAGTAATGC CAACCTAACT ATTAAACCA
 3901 AAGAAATTGAA ATTGACAGAA GACCTAAGTA TTTCAGGTTT CAATAAGCA
 3951 GAGATTACAG CCAAAGATGG TAGAGATTAA ACTATTGGCA ACAGTAAATGA
 4001 CGGTAACAGC GGTGCCGAAG CCAAAACAGT AACTTTAAC AATGTTAAAG

37/68

FIG. 7F.

4051 ATTCAAAAT CTCTGCTGAC GGTCAACAATG TGACACTAA TAGCAAAGTG
 4101 AAAACATCTA GCAGCAATGG CGGACCGTCAA AGCAATAGCG ACAACGATA
 4151 CGGCTTAACCT ATTACTGCAA AAAATGTAGA AGTAAACAAA GATATTACTT
 4201 CTCTCAAAAC AGTAAATATC ACCGGCTCGG AAAAGGTTAC CACCACAGCA
 4251 GGCTCGACCA TAAACGCAAAC AAATGGCAA GCAAGTATTA CAACCAAAAC
 4301 AGGTGATATC AGCGGTACGA TTTCCGGTAA CACGGTAAGT GTTAGCCGGA
 4351 CTGGTGATT AACCACTAAA TCCGGCTCAA AAATTGAAAGC GAAATCGGGT
 4401 GAGGCTAATG TAACAAGTGC AACAGGTACA ATTGGGGTA CAATTCCGG
 4451 TAATACGGTA AATGTTACGG CAAACCGCTGG CGATTAAACA GTTGGGAATG
 4501 GCGCAGAAAT TAATGGGACA GAAGGGAGCTG CAACCTTAAC CGCAACAGGG
 4551 AATAACCTTGA CTACTGAAGC CGGTTCTAGC ATCACTTCAA CTAAGGGTCA
 4601 GGTAGACCTC TTGGCTCAGA ATGGTAGGCAT CGCAGGAAGC ATTAATGCTG
 4651 CTAATGTGAC ATTAAATACT ACAGGCACCT TAACCACCGT GGCAAGGCTCG
 4701 GATATTAAAG CAAACCGGG CACCTTGGTT ATTAACGCAA AAGATGCTAA
 4751 GCTAAATGGT GATGCCATCAG GTGATAGTAC AGAAGTGAAT GCAGTCAAACG
 4801 ACTGGGGATT TGGTAGTGTG ACTGGGGCAA CCTCAAGCAG TGTGAATATC
 4851 ACTGGGGATT TAAACACAGT AAATGGGTTA AATATCATTT CGAAAGATGG

38 / 68

FIG. 7G.

4901 TAGAAACACT GTGGCCTTAA GAGGCAAGGA AATTGAGGTG AAATATATCC
 4951 AGCCAGGTGT AGCAAGTGT AAGAAAGTAA TTGAAGCGAA ACGCGTCCTT
 5001 GAAAAGTAA AAGATTTATC TGATGAAGAA AGAGAAACAT TAGCTAAACT
 5051 TGGTGTAAAGT GCTGTACGTT TTGTTGAGCC AAATAATACA ATTACAGTCA
 5101 ATACACAAA TGAAATTACA ACCAGACCGT CAAGTCAGT GATAATTCT
 5151 GAAGGTAAGG CGTGTTCCTC AAGTGGTAAT GGCGCACGAG TATGTACCAA
 5201 TGTTGCTGAC GATGGACAGC CGTAGTCAGT AATTGACAAG GTAGATTTC³⁹
 5251 TCCTGCAATG AAGTCATTT ATTTCGTAT TATTACTGT GTGGGTTAAA⁶⁸
 5301 GTtCAgTACG GGCTTTACCC ATCTTGTAAA AAATTACGGA GAATAACAATA
 5351 AAGTATTTT AACAGGTAT TATTATGAAA AATATAAAAA GCAGATTAAA
 5401 ACTCAGTCCA ATATCAGTAT TGCTTGGCCT GGCTTCTTC TCATTGTATG
 5451 CAGAAGAACG GTTTTAGTA AAAGGCTTTC AGTTATCTGG TGCACTTGAA
 5501 ACTTTAAGTGT AAGACGCCA ACTGTTGTGA GCAAAATCTT TATCTAAATA
 5551 CCAAGGCTCG CAAACTTAA CAAACCTAAA AACAGCACAG CTTGAATTAC
 5601 AGGCCTGTGCT AGATAAGATT GAGCCAAATA AATTGATGT GATATTGCCG
 5651 CAAACAAACCA TTACGGATGG CAATATCATG TTTGAGCTAG TCTCGAAATC

FIG. 7H.

5701 AGCCGGCAGAA AGCCAAGTTT TTATATAAGGC GAGCCAGGGT TATAGTGAAG
 5751 AAAATATCGC TCGTAGCCTG CCATCTTGA ACAAGGAAA AGTGTATGAA
 5801 GATGGTCGTC AGTGGTCGA TTTGGCTGAA TTTAATATGG CAAAGAAAA
 5851 CCCGCTTAAG GTTACCCGTG TACATTAACGA ACTAAACCTT AAAAACAAAA
 5901 CCTCTAATT GATAATTGCG GGCTTCTCGC CTTTTGGTAA AACGGCTAGC
 5951 TTTATTCTT ATGATAATT CGGGCGAGA GAGTTAACT ACCAACGTGT
 6001 AAGCTTGGGT TTTGTTAATG CCAATTAAAC TGGTCATGAT GATGTGTTAA
 6151 TTATACCAGT ATGAGTTATG CTGATTCTAA TGATATCGAC GGCTTACCAA
 6201 GTGCGATTAA TCGTAAATT TCAAAGGTC AATCTATCTC TGCGAATCTG
 6251 AAATGGAGTT ATTATCTCCC AACATTAAAC CTTGGCATGG AAGACCAATT
 6301 TAAAATTAAAT TTAGGCTACA ACTACCGCCA TATTAATCAA ACCTCCGGGT
 6351 TAAATCGCTT GGGTGAACG AAGAAAAAT TTGCAGTATC AGGGCTAAGT
 6401 GCAGGCATTG ATGGACATAT CCATTTACC CCTAAACAA TCTTTAATAT
 6451 TGATTAACT CATCATTATT ACGCGAGTAA ATTACCAAGGC TCTTTGAA
 6501 TGGAGCCAT TGGCGAAACA TTAAATCGCA GCTATCACAT TAGCACAGCC
 6551 AGTTTAGGGT TGAGTCAAGA GTTGTGCTAA GGTTGGCATT TTAGCAGTCA
 6601 ATTATCAGGT CAATTACTC TACAAGATAT TAGCAGTATA GATTATTCT

FIG. 7I.

6651 CTGTAACAGG TACTTATGGC GTCAGGGCT TAAATAACGG CGGTGCAAGT
 6701 GGTGAGGGCG GTCTTGTATG GCGTAAATGAA TTAAGTATGC CAAATAACAC
 6751 CCGCTTCCAA ATCAGCCCTT ATGCCTTTA TGATGCCAGGT CAGTCCGTT
 6801 ATAATAGCGA AAATGCTAAA ACTTACGGCG AAGATATGCA CACGGTATCC
 6851 TCTGCGGGTT TAGGCATTA AACCTCTCCT ACACAAACT TAAGCCTAGA
 6901 TGCTTTGTT GCTCGTGGCT TTGCAAATGC CAATAGTGC AATTGAAATG
 6951 GCAACAAAAA ACGCACAGC TCACCTACAA CCTCTGGG GAGATTAACA 41
 7001 TTCAGTTCT AACCCCTGAAA TTTAATCAAC TGGTAAGCGT TCCGCCCTACC 68
 7051 AGTTATAAC TATATGCTT ACCGCCAAT TTACAGTCTA TAGGCAACCC
 7101 TGTTTTAC CTTATATTC AAATAACAA GCTAAGCTGA GCTAAGCAA
 7151 CCAAGCAAC TCAAGCAAGC CAAGTAATAC TAAAAAACAA ATTATATGA
 7201 TAAACTAAAG TATACTCCAT GCCATGGCGA TACAAGGGAT TTAATAATAT
 7251 GACAAAGAA AATTGCAAA ACGCTCCTCA AGATGCGACC GCTTTACTTG
 7301 CGGAATTAAAG CAACAATCAA ACTCCCCCTGC GAATATTAA ACAACCAACCGC
 7351 AAGCCCAGCC TATTACGCTT GGAACAAACAT ATCCAAAAA AAGATTATGA
 7401 GTTTGCTTGT CGTGAATTAA TGGTGATTCT GGAAAAATG GACGCTAATT

FIG. 7J.

7451 TTGGAGGGGT TCACGATATT GAATTGACG CACCGCTCA GCTGGCATAT
 7501 CTACCCGAAA AATTACTAAT TTATTGGCC ACTCGTCTCG CTAATGCAAT
 7551 TACAACACTC TTTCCGACC CCGAATTGGC AATTCTGAA GAAGGGGGGT
 7601 TAAAGATGAT TAGCCTGCAA CGCTGGTTGA CGCTGATT TTGCTCTTCC
 7651 CCCTACGTTA ACGCAGACCA TATTCTCAAT AAATATAATA TCAACCCAGA
 7701 TTCCGAAGGT GGCTTTCAT TAGCAACAGA CAACTCTCT ATTGCTAAAT
 7751 TCTGTATT TTACTTACCC GAATCCAATG TCAATATGAG TTTAGATGCG 42 / 68
 7801 TTATGGCAG GGAATCAACAA ACTTTGGCT TCATTGTT TTGCGTTGCA
 7851 GTCTTCACGT TTATGGTA CCGCATTCTGC GTTCATAAA AGAGCGGTGG
 7901 TTTTACAGTG GTTTCCTAAA AAACCTGCCG AAATTGCTAA TTTAGATGAA
 7951 TTGCCTGCAA ATATCCTTCA TGATGTATAT ATGCACTGCA GTTATGATT
 8001 AGCAAAAAC AAGCACGATG TTAAGCGTCC ATTAAACGAA CTTGTCCGCA
 8051 AGCATATCCT CACGCAAGGA TGGCAAGACC GCTACCTTA CACCTTAGGT
 8101 AAAAGGACG GCAAACCTGT GATGATGGTA CTGCTTGAAC ATTAAATTTC
 8151 GGGACATTGG ATTATCGTA CACATTCAAC TTCAATGATT GCTGCTCGAG
 8201 AAAAATTCTA TTAGTGGGC TAGGCCATG AGGGCGTTGA TAAATAGGT

FIG. 7K.

8251 CGAGAAGTGT TTGACCGAGTT CTTTGAATC AGTAGCAATA ATATAATGGA
 8301 GAGACTGTGTT TTTATCCGTA AACAGTGCAG AACTTTCCAA CCCGCAGTGT
 8351 TCTATATGCC AAGCATTGGC ATGGATATT CCACGATT TTGAGGCAAC
 8401 ACTCGGCTTG CCCCTATTCA AGCTGTAGCC CTGGGTCACTC CTGCCCACTAC
 8451 GCATTCTGAA TTTATTGATT ATGTCATCGT AGAAGATGAT TATGTGGCA
 8501 GTGAAGGATTG TTTCAGCGAA ACCCTTTAC GCTTACCCAA AGATGCCCTA
 8551 CCTTATGTAC CTTCTGCACT CGCCCCACAA AAAGTGGATT ATGTA
 8601 GGAAAACCTT GAAGTAGTCA ATATCGGTAT TGCCGCTACC ACAATGAAAT 43 / 60
 8651 TAAACCTGAA ATTGTTGCTA ACATTGCAAG AAATCAGAGA TAAAGCTAAA
 8701 GTCAAAATAC ATTTCATT CGCACTTGGAA CAATCAACAG GCTTGACACA
 8751 CCCTTATGTC AAATGGTTA TCGAAAGCTA TTTAGGTGAC GATGCCACTG
 8801 CACATCCCCA CGCACCTTAT CACGATTATC TGGCAATT ATTGCTGATT
 8851 GATATGCTAC TAAATCCGTT TCCTTCCGGT AATACTAACG GCATAATTGA
 8901 TATGGTTACA TTAGGTTAG TTGGTGTATG CAAACGGGG GATGAAGTAC
 8951 ATGAACATAT TGATGAAGGT CTGTTAAC GCTTAGGACT ACCAGAATGG
 9001 CTGATAGCCG ACACACGAGA AACATATATT GAATGTCCT TGCGTCTAGC
 9051 AGAAAACCATT CAAGAACGCC TTGAACCTCCG TCGTTACATC ATAGAAAACA

FIG. 7L.

9101 ACGGCTTACA AAAGCTTTTT ACAGGGCGACC CTCGTCCATT GGGCAAAATA
9151 CTGCTTAAGA AAACAAATGA ATGGAAGCGG AAGCACTTGA GTAAAAAATA
9201 ACGGTTTTT AAAGTAAAG TGCGGTTAAT TTTCAAAGCG TTTAAAAAC
9251 CTCTCAAAA TCAACCGCAC TTTTATCTT ATAACGATCC CGCACGGCTGA
9301 CAGTTATCA GCCTCCCCGC ATA-AAACTCC GCCTTTTCATG GGGGAGATT
9351 TAGCCAAAC TGGCAGAAAT TAAAGGCTAA AATCACCAA TTGCACCCACA
9401 AAATCACCAA TACCCACAA AAA

FIG. 8A.

1 GATCAATCTG GGCGATATT TTGCCAAAGG TGGTAACATT AATGTCGGCG
 51 CTGCCACTAT TCGCAATAAA GGTAAACTTT CTGCCGACTC TGTAAAGCAA
 101 GATAAAAGTG GTAACATTGT TCTCTCTGCC AAAGAAGGTG AAGCGGAAAT
 151 TGGCGGTGTA ATTTCGGCTC AAAATCAGCA AGCCAAAGGT GGTAAAGTTGA
 201 TGATTACAGG CGATAAAAGTT ACATTGAAAA CGGGTGCAGT TATCGACCTT
 251 TCGGGTAAAG AAGGGGGAGA AACTTATCTT GCCGGTGACG AGCGTGGCGA
 301 AGGTAAAAC GGCATTCAAT TAGCAAAGAA AACCACTTTA GAAAAGGCT 45 /
 351 CAACAATTAA TGTGTCAAGGT AAAGAAAAAG GTGGGGCGGC TATTGTATGG 60
 401 GGGGATATTG CGTTAATTGA CGGCAATATT AATGCCAAG GTAAAGATAT
 451 CGCTAAACT GGTGGTTTG TGGAGACGTC GGGCATTAC TTATCCATTG
 501 ATGATAACGC AATTTGTTAAA ACAAAAGAAT GGCTACTAGA CCCAGAGAAT
 551 GTGACTATTG AAGCTCCTTC CGCTTCTCGC GTCGAGCTGG GTGCCGATAG
 601 GAATTCCAC TCGGCAGAGG TGATAAAAGT GACCCCTAAAA AAAATAACA
 651 CCTCCTTGAC AACACTAAC AATACAAACCA TTTCAAATCT TCTGAAAAGT
 701 GCCCACGTGG TGAACATAAC GGCAAGGAGA AAACTTACCG TTAATAGCTC
 751 TATCAGTATA GAAAGAGGCT CCCCACTTAAT TCTCCACAGT GAAGGGTCAGG

FIG. 8B.

801 GCGGTCAAAGG TGTTCAGATT GATAAAGATA TTACTTCTGA AGGGGAAAT
 851 TTAACCATTT ATTCTGGGG ATGGGTGAT GTTCATAAAA ATATTACGCT
 901 TGGTAGCGGC TTTTAAACA TCACAACTAA AGAAGGAGAT ATCGCCTTCG
 951 AAGACAAAGTC TGGACCGGAAC AACCTAACCA TTACAGCCC AGGGACCATC
 1001 ACCTCAGGTA ATAGTAACGG CTTTAGATT ACAAACGGTCT CTCTAAACAG
 1051 CCTTGGCGGA AAGCTGAGCT TTACTGACAG CAGAGGGAC AGAGGTAGAA
 1101 GAACTAAGGG TAATATCTCA AACAAATTG ACGGAACGTT AAACATTCC
 1151 GGAACGTAG ATATCTCAAT GAAAGCACCC AAAGTCAGCT GGTTTACAG 46 / 68
 1201 AGACAAAGGA CGCACCTACT GGAAACGTAAC CACTTTAAT GTTACCTCGG
 1251 GTAGTAAATT TAACCTCTCC ATTGACAGCA CAGGAAGTGG CTCAACAGGT
 1301 CCAAGCATAAC GCAATGCAGA ATAAATGGC ATAACATTAA ATAAGGCCAC
 1351 TTTTAATATC GCACAAGGCT CAACAGCTAA CTTTAGCATE AAGGCATCAA
 1401 TAATGCCCTT TAAGAGTAAC GCTAACTACG CATTATTAA TGAAGATATT
 1451 TCAGTCTCAG GGGGGGTAG CGTTAATTTC AAACTTAACG CCTCATCTAG
 1501 CAAACATACAA ACCCTGGCC TAATTATAAA ATCTCAAAAC TTTAATGTCT
 1551 CAGGAGGGTC AACTTTAAAT CTCAGGGCTG AAGGTTCAAC AGAAACCGCT
 1601 TTTTCAATAG AAAATGATT AAACCTAACC GCCACCGGTG GCAATATAAC

FIG. 8C.

1651 AATCAGACAA GTCGAGGGTA CCGATTCACG CGTCAACAAA GGTGTGGCAG
 1701 CCAAAAAAA CATAACTTTT AAAGGGGGTA ATATCACCTT CGGCTCTCAA
 1751 AAAGCCACAA CAGAAATCAA AGGCAATGTT ACCATCAATA AAAACACTAA
 1801 CGCTACTCTT CGTGGTGCAG ATTGCGGA AAACAAATCG CCTTTAAATA
 1851 TAGCAGGAAA TGTATTAAAT AATGGCAACC TTACCACTGC CGGCTCCATT
 1901 ATCAATATAG CCGGAATCT TACTGTTCA AAAGGGGCTA ACCTTCAAGC
 1951 TATAACAAAT TACACTTTA ATGTAGCCGG CTCATTGAC ACAATGGCG
 2001 CTTCAAAACAT TTCCATTGCC AGAGGGGGG CTAAATTAA AGATATCAAT
 2051 AACACCAAGTA GCTTAATAT TACCAACAC TCTGATACCA CTTACCGCAC
 2101 CATTATAAA GGCAATATAT CCAACAAATC AGGTGATTG AATATTATTG
 2151 ATAAAAAAAG CGACGCTGAA ATCCAAATTG GGGCAATAT CTCACAAAAA
 2201 GAAGGCAATC TCACAATTTC TTCTGATAAA GTAAATATA CCAATCAGAT
 2251 AACAAATCAA GCAGGGCTTG AAGGGGGCG TTCTGATTCA AGTGAGGCAG
 2301 AAAATGCTAA CCTAACTATT CAAACCAAAG AGTTAAATT GGCAGGGAGAC
 2351 CTAATATT CAGGCTTTAA TAAAGCAGAA ATTACAGCTA AAATGGCAG
 2401 TGATTAACT ATTGGCAATG CTAGGGTGG TAATGCTGAT GCTAAAAAAG

47 / 68

FIG. 8.D.

2451	TGACTTTGAA	CAAGGTTAAA	GATTCAAAAA	TCTCGACTGAA	CGGTCACAAAT
2501	GTAAACACTAA	ATAGCGGAAGT	GAAAACGTCT	AATGGTAGTA	GCAATGCTGG
2551	TAATGATAAC	AGCACCGGGTT	TAACCATTTC	CGCAAAAGAT	GTAACGGTAA
2601	ACAATAACGT	TACCTCCCAC	AAGACAAATAA	ATATCTCTGC	CGCAGCAGGA
2651	AATGTAACAA	CCAAAGGAAGG	CACAACTATC	AATGCAACCA	CAGGCAGCGT
2701	GGAAAGTAACCT	GCTCAAAATG	GTACAATTAA	AGGCAACAT	ACCTCGCAA
2751	ATGTAACAGT	GACAGGCAACA	GAAAATCTTG	TTACCACAGA	GAATGCTGTC
2801	ATTAAATGCAA	CCAGCGGCAC	AGTAAACATT	AGTACAAAAA	CAGGGATAT
2851	TAAAGGTTGGA	ATTGAATCAA	CTTCCGGTAA	TGTAAATATT	ACAGCGAGCG
2901	GCAATACACT	TAAGGTTAAGT	AATATCACTG	GTCAAGAGATGT	AACAGTAACA
2951	GCGGATGCAG	GAGGCCCTGAC	AACTACAGCA	GGCTCAACCA	TTAGTGGGAC
3001	AACAGGCAAT	GCAAATAATTA	CAACCAAAAC	AGGTGATATC	AACGGTAAAG
3051	TTGAATCCAG	CTCCGGCTCT	GTAACACTTG	TTGCAACTG	AGCAAACCTTT
3101	GCTGTAGGTA	ATATTCAGG	TAACACTGTT	ACTATTACTG	CGGATAGCGG
3151	TAATTAACC	TCCACAGTAG	GTTCTACAAT	TAATGGGACT	AATACTGTAA
3201	CCACCTCAAG	CCAATCAGGC	GATATTGAAG	GTACAAATTTC	TGGTAATAACA
3251	GTAATATGTTA	CAGGCAAGCAC	TGGTGATTAA	ACTATGGAA	ATAGTGCAA

FIG. 8E.

3301 AGTTGAAGCG AAAAATGGAG CTGCAACCTT AACTGCTGAA TCAGGCAAAT
 3351 TAACCACCCA AACAGGCCA AGCATTAACCT CAAGCAATGG TCAGACAACT
 3401 CTTACAGCCA AGGATAGCAG TATCGCAGGA AACATTAATG CTTGCTAATGT
 3451 GACGTTAAAT ACCACAGGCA CTTTAACATC TACAGGGAT TCAAAGGATA
 3501 ACGCAACCAG TGGTACCTTA ACAATCAATG CAAAGATGC CAAATTAGAT
 3551 GGTGCTGCAT CAGGTGACCG CACAGTAGTA AATGCAACTA ACGCAAGTGG
 3601 CTCTGGTAAAC GTGACTGCGA AACCTCAAG CAGCGTGAAT ATCACCGGGG
 3651 ATTTAACAC ATAATGGG TTAATATCA TTTCGGAAA TGGTAGAAAC
 3701 ACTGTGGCT TAAGAGGCAA GGAATTGAT GTGAAATATA TCCAACCAGG
 3751 TGTAGCAAGC GTAGAAGAGG TAATTGAAGC GAAACGCGTC CTTGAGAAGG
 3801 TAAAAGATT ATCTGATGAA GAAAGAGAAA CACTAGCCAA ACTTGGTGTGA
 3851 AGTGCTGTAC GTTTCGTTGA GCCAAATAAT GCCATTACGG TTAATACACA
 3901 AAACGAGTT ACAACCAAC CATCAAGTCA AGTGACAATT TCTGAAGGTA
 3951 AGGCCGTGTT CTCAAAGTGGT ATGGCCAC GAGTATGTAC CAATGTTGCT
 4001 GACGATGGAC AGCAGTAGTC AGTAATTGAC AAGGTAGATT TCATCCCTGCA
 4051 ATGAAAGTCAT TTTATTTAC TATTATTTAC TGTGTGGTT AAAGTTCACT

50/68

FIG. 8F.

4101 ACGGGCTTA CCCACCTTGT AAAAATTAC GAAAATACA ATAAAGTATT
4151 TTTAACAGGT TATTATTATG AAAAACATAA AAAGCAGATT AAAACTCAGT
4201 GCAATATCAA TATTGCTTGG CTTGGCTTCT TCATGGACGT ATGCAGAAGA
4251 AGCGTTTTA GTAAAAGGCT TTCAGTTATC TGGCGCG

FIG. 9A.

1 GGGAAATGAGC GTCGTACACCG GTACAGCAAC CATGCAAAGTA GACGGCAATA
 51 AAACCACTAT CCGTAATAGC GTCAATGCTA TCATCAAATTG GAAACAAATT
 101 AACATTGACC AAAATGAAAT GGAGGCAGTTT TTACAAGAAA GCAGCAACTC
 151 TGCCCGTTTTC ACCCGTGTAA CACCTGACCA AATCTCCCAA TAAAAGGGA
 201 TTTTAGATTTC TAACGGACAA GTCTTTTAA TCAACCCAAA TGGTATCACCA
 251 ATAGGTAAAG ACGCAATTAT TAACACTAAT GGCTTACTG CTTCTACGCT
 301 AGACATTCTT AACGAAAACA TCAAGGGCGC TAATTTCACCC CTTGAGCAA
 351 CCAAGGATAA AGCACTCGCT GAAATCGTGA ATCACGGTT ATTACCGTT
 401 GGTAAAGAGCG GTAGCGTAAA CCTTATTGGT GGCAAAGTGA AAAACGAGGG
 451 CGTGATTAGC GTAAATGGCG GTAGTATTTC TTACTTGCA GGGCAAAAA
 501 TCACCATCAG CGATATAATA AATCCAACCA TCACCTACAG CATTGCTGCA
 551 CCTGAAACG AAGCGATCAA TCTGGCGAT ATTTTGCCA AAGGTGGTAA
 601 CATTAAATGTC CGCGCTGCCA CTATTGCCA TAAAGGTAAA CTTTCTGCCG
 651 ACTCTGTAAAG CAAAGATAAA AGTGGTAACA TTGGTTCTC TGCCAAAGAA
 701 GGTGAAGCGG AAATTGGCGG TGTAATTTC GCTCAAAATC AGCAAGCAA
 751 AGGTGGTAAG TTGATGATTA CAGGTGATAA AGTCACATTA AAAACAGGTG

FIG. 9B.

801 CAGTTATCGA CCTTTCAGGT AAAGAAGGG GAGAGACTTA TCTTGGCGGT
 851 GATGAGCGTG GCGAAGGTAA AAATGGTATT CAATTAGCGA AGAAAACCTC
 901 TTTAGAAAAA GGCTCGACAA TTATGTATC AGGCAAAGAA AAAGGGGGC
 951 GCGCTATTGT ATGGGGGAT ATTGCATTAA TTAATGGTAA CATTATGCT
 1001 CAAGGTAGCG ATATTGCTAA AACTGGCGC TTTGTGAAA CATCAGGACA
 1051 TGACTTATCC ATTGGTGTATC ATGGTGTATT TGACGCTAAA GAGTGGTTAT
 1101 TAGACCCAGA TGATGTGTCC ATTGAAACTC TTACATCTGG ACGCAATAAT
 1151 ACCGGCGAAA ACCAAGGATA TACAAACAGGA GATGGGACTA AAGAGTCACC 52 / 68
 1201 TAAAGGTAAT AGTATTCTA AACCTACATT AACAAACTCA ACTCTTGAGC
 1251 AAATCCTAAC AAGAGGTTCT TATGTTAATA TCACTGCTAA TAATAGAATT
 1301 TATGTTAATA GCTCCCATCAA CTTATCTAAT GGCAGTTAA CACTTCACAC
 1351 TAAACGAGAT GGAGTTAAA TTAAACGGTGA TATTACCTCA AACGAAAATG
 1401 GTAATTAAAC CATTAAAGCA GGCTCTGGG TTGATGTTCA TAAAACATC
 1451 ACGCTTGGTA CGGGTTTTT GAATATTGTC GCTGGGGATT CTGTAGCTT
 1501 TGAGAGAGAG GGGGATAAAG CACGTAACGC AACAGATGCT CAAATACCG
 1551 CACAAGGGAC GATAACCGTC AATAAAGATG ATAAACAAATT TAGATCAAT
 1601 AATGTTATCTA TAAACGGGAC GGGCAAGGGT TAAAGTTA TTGCAAATCA

FIG. 9C.

1651 AAATAATTCT ACTCATAAAAT TTGATGGCGA AATTAACATA TCTGGAAATAG
 1701 TAACAAATTAA CCAAACCCACG AAAAAAGATG TAAATAACTG GAATGCATCA
 1751 AAAGACTCTT ACTGGAAATGT TTCTTCTCTT ACTTTGAAATA CGGTGGAAAA
 1801 ATTACCTTT ATAATTCG TTGATAGCGG CTCAAATICC CAAGATTGAA
 1851 GGTCACTCACG TAGAAGTTT GCAGGGCGTAC ATTAAACCGG CATCGGAGGC
 1901 AAAACAAACT TCAACATCGG AGCTAACGCA AAAGCCTTAT TAAATTAAGG 53/68
 1951 ACCAAACGCC GCTACAGACC CAAAAAAAGA ATTACCTATT ACTTTAACG
 2001 CCAACATTAC AGCTACCGGT AACAGTGATA GCTCTGTGAT GTTTGACATA
 2051 CACGCCAATC TTACCTCTAG AGCTGCCGGC ATAAACATGG ATTCAATTAA
 2101 CATTACCGGC GGGCTTGACT TTTCATAAC ATCCCATAAT CGCAATAGTA
 2151 ATGCTTTGAA AATCAAAAAA GACTTAACTA TAAATGCAAC TGGCTCGAAT
 2201 TTTAGTCTTA AGCAAACGAA AGATTCTTT TATAATGAAT ACAGCAAACA
 2251 CGCCATTAAAC TCAAGTCATA ATCTAACCAT TCTTGGGGC AATGTCACTC
 2301 TAGGTGGGA AAATTCAAGC AGTAGCATTAA CGGGCAATAT CAATATCACC
 2351 AATAAAGCAA ATGTTACATT ACAAGCTGAC ACCAGCAACA GCAACACAGG
 2401 CTTGAAGAAA AGAAACTCTAA CTCTTGGCAA TATATCTGTT GAGGGGAATT

FIG. 9D.

2451 TAAGCCTAAC TGGTCAAAT GCAAACATG TCGGCAATCT TTCTATTGCA
 2501 GAAGATTCCA CATTAAAGG AGAACCCAGT GACAACCTAA ACATCACCGG
 2551 CACCTTACCA AACAACGGTA CCGCCAACAT TAATATAAA CAAGGAGTGG
 2601 TAAAACCTCCA AGGCCATATT ATCAAATAAG GTGGTTAAA TATCACTACT
 2651 AACGCCCTCAG GCACTCAAAA AACCAATT ATT AACGGAAATA TAACTAACGA
 2701 AAAAGGGGAC TTAAACATCA AGAATTTAA AGCCGACGCC GAAATCCAA
 2751 TTGGGGCAA TATCTCACAA AAAGAAGGCA ATCTCACAAAT TTCTTCTGAT 54
 2801 AAAGTAAATA TTACCAATCA GATAACAATC AAAGCAGGGCG TTGAAGGGGG
 2851 GCGTTCTGAT TCAAGTGAGG CAGAAAATGC TAACTTAACCT ATTCAAACCA 68
 2901 AAGAGTTAAA ATTGGCAGGA GACCTAAATA TTTCAGGCCCT TAATAAAGCA
 2951 GAAATTACAG CTAAAATGG CAGTGATTAA ACTATTGGCA ATGCTAGCGG
 3001 TGGTAATGCT GATGCTAAA AAGTGACTTT TGACAAGGTT AAAGATTCAA
 3051 AAATCTCGAC TGACGGTCAC AATGTAACAC TAAATAGCGA AGTGAACCG
 3101 TCTAATGGTA GTAGGAATGTC TGGTAATGAT AACAGCACCG GTTTAACCAT
 3151 TTCCGCAAAA GATGTAACGG TAAACAATAA CGTTACCTCC CACAAGACAA
 3201 TAAATATCTC TGCCGAGCA GGAAATGTA CACCAAAGA AGGCACAACT
 3251 ATCAATGCCA CCACAGGCAG CGTGGAAAGTA ACTGCTCAA ATGGTACAAT

FIG. 9E.

3301 TAAAGGCAAC ATTACCTCGC AAAATGTAAC AGTGACAGCA ACAGAAAATC
 3351 TTGTTACCAAC AGAGAATGGCT GTCATTAATG CAACCAGCGG CACAGTAAAC
 3401 ATTAGTACAA AACAGGGAA TATTAAGGT GGAATTGAAAT CAACTTCCGG
 3451 TAATGTAAT ATTACAGCGA GGGCAATAAC ACTTAAGGTA AGTAATATCA
 3501 CTGGTCAAGA TGTAAACAGTA ACAGGGGATG CAGGAGCCCTT GACAACCTACA
 3551 GCAGGGCTCAA CCATTAGTGC GACAACAGGC AATGCAAATA TTACACCAA
 3601 AACAGGTGAT ATCAACGGTA AAGTTGAATC CAGCTCCGGC TCTGTAACAC 55
 3651 TTGTTGCAAC TGGAGCAACT CTTGCTGTAG GTAAATTTTC AGGTAACACT 68
 3701 GTTACTATTAA CTGGGGATAG CGGTAAAATTAA ACCTTCCACAG TAGGTTCTAC
 3751 ATTAAATGGG ACTTAATAGTG TAACCACCTC AAGCCAATCA GGGGATATTG
 3801 AAGGTACAAT TTCTGGTAAT ACAGTAAATG TTACAGCAAG CACTGGTGAT
 3851 TTAACTATTG GAAATAGTGC AAAAGTTGAA GCGAAAAATG GAGCTGCAAC
 3901 CTTAACTGCT GAATCAGGCA ATTAAACCAC CCAAACAGGC TCTAGGATTA
 3951 CCTCAAGCAA TGGTCAGACA ACTCTTACAG CCAAGGATAG CAGTATCGCA
 4001 GGAAACATTA ATGCTGCTAA TGTGACGTTA AATACCACAG GCACTTTAAC
 4051 TACTACAGGG GATTCAAAGA TTAACGGCAAC CAGTGGTACC TAAACAATCA

FIG. 9F.

4101	ATGCCAAAGA	TGCCAAATTAA	GATGGTGCTG	CATCAGGTGA	CCGGCACAGTA
4151	GTAATGCAA	CTAACGCCAAG	TGGCTCTGGT	AACGTGACTG	CGAAAACCTC
4201	AAGCAGCGTG	AATATCACCG	GGGATTAAA	CACAATAAAT	GGGTTAAATA
4251	TCATTTCGGA	AAATGGTAGA	AACACTGTGC	GCTTAAGAGG	CAAGGAAATT
4301	GATGTGAAAT	ATATCCAACC	AGGTGTAGCA	AGCGTAGAAG	AGGTAATTGTA
4351	AGCGAACCG	GTCCTTGAGA	AGGTAAAAGA	TTTATCTGAT	GAAGAAAGAG
4401	AAACACTAGC	CAAACTTGGT	GTAAGTGCCTG	TACGTTTCGTT	TGAGGCCAAT
4451	AATGCCATTAA	CGGTTAAATAC	ACAAAACCGAG	TTTACAAACCA	AACCATCAAG
4501	TCAAGTGACA	ATTTCCTGAAG	GTAAGGCCGTG	TTTCTCAAGT	GGTAATGGCG
4551	CACGAGTATG	TACCAATGTT	GCTGACGGATG	GACAGCAGTA	GTCAGTAATT
4601	GACAAGGTAG	ATTTCATCCT	GCAATGAAGT	CATTTTATT	TCGTATTATT
4651	TACTGTGTGG	GTTAAAGTTC	AGTACGGGCT	TTACCCACCT	TGTAAAAAAT
4701	TA				

FIG. 10A. COMPARISON OF DERIVED AMINO ACID SEQUENCE

FIG. 10B.

Hmw1.com NWKQFN1DQN EMVQFLQENN NSAVFNRVTS NQISQLKGIL DSNQVFLIN
 Hmw2.com NWKQFN1DQN EMVQFLQENN NSAVFNRVTS NQISQLKGIL DSNQVFLIN

151	200	250	300
Hmw3.com
Hmw4.com PNGITIGKDA IINTNGFTAS TLDISNENIK ARNFTLEQTK DKALAEIVNH
Hmw1.com PNGITIGKDA IINTNGFTAS TLDISNENIK ARNFTLEQTK DKALAEIVNH 5
Hmw2.com PNGITIGKDA IINTNGFTAS TLDISNENIK ARNFTLEQTK DKALAEIVNH 68
.....			
201	250	251	300
Hmw3.com
Hmw4.com GLITVGKDGS VNLIGGKVKN EGVISVNGGS ISLLAGQKIT ISDIINPTIT
Hmw1.com GLITVGKDGS VNLIGGKVKN EGVISVNGGS ISLLAGQKIT ISDIINPTIT
Hmw2.com GLITVGKDGS VNLIGGKVKN EGVISVNGGS ISLLAGQKIT ISDIINPTIT
.....			
Hmw3.com	INLGDIFAKG GNINVRAATI RNKGKLSADS VSKDKSGNIV		

FIG. 10C.

Hmw4.com	Y SIAAPNEA	INLGDIFAKG	GNINVRAATI	RNKGKLLSADS	VSKDKSGNIV
Hmw1.com	Y SIAAPNEA	VNLGDIFAKG	GNINVRAATI	RNKGKLLSADS	VSKDKSGNIV
Hmw2.com	Y SIAAPNEA	VNLGDIFAKG	GNINVRAATI	RNKGKLLSADS	VSKDKSGNIV

301

Hmw3.com	LSAKEGEAEI	GGVISAQNQQ	AKGGKLMITG	DKVTLKTGAV	IDLSGKEGG 59
Hmw4.com	LSAKEGEAEI	GGVISAQNQQ	AKGGKLMITG	DKVTLKTGAV	IDLSGKEGG 68
Hmw1.com	LSAKEGEAEI	GGVISAQNQQ	AKGGKLMITG	DKVTLKTGAV	IDLSGKEGG
Hmw2.com	LSAKEGEAEI	GGVISAQNQQ	AKGGKLMITG	DKVTLKTGAV	IDLSGKEGG

350

Hmw3.com	TYLGDERGE	GKNGIQLAKK	TTLEKGSTIN	VSGKEKGGRA	IVWGDIALID
Hmw4.com	TYLGDERGE	GKNGIQLAKK	TTLEKGSTIN	VSGKEKGGRA	IVWGDIALID
Hmw1.com	TYLGDERGE	GKNGIQLAKK	TTLEKGSTIN	VSGKEKGGRA	IVWGDIALID
Hmw2.com	TYLGDERGE	GKNGIQLAKK	TTLEKGSTIN	VSGKEKGGRA	IVWGDIALID

350

400

FIG. 10D.

401

Hmw3.com	GNINAQGK.D	IAKTGGFVET	SGHYLSIIDDN	AIVKTKEWLL	DOPENVTIEAP
Hmw4.com	GNINAQGGS.D	IAKTGGFVET	SGHDLSIGDD	VIVDAKEWLL	DPPDVSIETL
Hmw1.com	GNINAQGSGD	IAKTGGFVET	SGHDLFIKDN	AIVDAKEWLL	DPDNVТИNAE
Hmw2.com	GNINAQGSGD	IAKTGGFVET	SGHYLSIESN	AIVKTKEWLL	DPPDVVTIEAE

450

Hmw3.com	SASRVELGAD	RNSHSAEVIK	VTLKKNNNTSL	TTLTNTTISN	LLKSAHVVN1
Hmw4.com	TSGRNMNTGEN	QGYTTGDGTK	ESPKGNNSISK	PTLTNSTLEQ	ILRRGSYVN1
Hmw1.com	TAGRSNTSED	DEYTGSGNSA	STPKRNKE.K	TTLTNTTLES	ILKKGTFVN1
Hmw2.com	DPLRNMNTGIN	DEFPTGTGEA	SDPKKNSELK	TTLTNTTISN	YLKNAWTMNI

450

60 / 68

Hmw3.com	TARRKLTVNS	SISIERGSHL	ILHSEGQQGQ	GVQIDKDITS	.E...GGNLT
Hmw4.com	TANNRIYVNS	SINLNSNGS.L	TLHTK...RD	GVKINGDITS	NE...NGNLT
Hmw1.com	TANQRIYVNS	SINL.SNGSL	TLWSEGRSGG	GVEINNDIT	GDDTRGANLT
Hmw2.com	TASRKLTVNS	SINGSNGL	ILHSKGQRGG	GVQIDGDT.	...SKGGNLT

500

501

Hmw3.com	TARRKLTVNS	SISIERGSHL	ILHSEGQQGQ	GVQIDKDITS	.E...GGNLT
Hmw4.com	TANNRIYVNS	SINLNSNGS.L	TLHTK...RD	GVKINGDITS	NE...NGNLT
Hmw1.com	TANQRIYVNS	SINL.SNGSL	TLWSEGRSGG	GVEINNDIT	GDDTRGANLT
Hmw2.com	TASRKLTVNS	SINGSNGL	ILHSKGQRGG	GVQIDGDT.	...SKGGNLT

FIG. 10E.

551

Hmw3com IYSGGWVDVH KNITLGS.GF LNITKEGDI AFEDKSGR. . . NNLTTITAQ
 Hmw4com IKAGSWVDVH KNITLGT.GF LNIVAGDS.V AFEREGDKAR NATDAQITAQ
 Hmw1com IYSGGWVDVH KNIISLGAQGN INITAKQD.I AFEKGNSNQV. ITGQ
 Hmw2com IYSGGWVDVH KNITLTD.QGF LNITA.AS.V AFEGGNNKAR DANNLTITAQ

600

601 61/68
 Hmw3com GTITSG.NSN GFRFNNVSLN SLGGKLSFTD SREDRGRRTK GNISNKFDGT
 Hmw4com GTITVNKDDK QFRFNNVSLN GTGKGKLFIA NQN. NFTHKFDGE
 Hmw1com GTIT.SGNQK GFRFNNVSLN GTGSGLQFTT KRTN. . . . K YAITNKFEGT
 Hmw2com GTVTITGECK DFRANNVSLN GTGKGGLNIIS SVNN. LTHNLSGT

650

651 700
 Hmw3com LNISGTVDIS MKAPKVSWFY RD.KGRTYWN VTTLMNVTSGS KFNLSIDSTG
 Hmw4com INISGIVTIN QTTKKDVKYW NA.SKDSYWN VSSLTLNTVQ KFTF.IKFVD
 Hmw1com LNISGKVNIS MVLPKNESGY DKFKGRTYWN LTSLMNVSESG EFNLTIDSRG

FIG. 10F.

Hmw2com INISGNITIN QTTRKNTSYW QTSHD. SHWN VSALNLETGA NFTF.IKYIS

701

750

Hmw3com SGSTG...PS IRNA.ELNG ITFN....KA TNIAQGSTA NFSIKASIMP
 Hmw4com SGNS...QD LRSSRRSFAG VFNGIGGKT NFNIGANAKA LFKLKPNAAT
 Hmw1com SDSAGTLTQ.PYNLNG ISFN...KDT TFNVERNARV NFDIKAPIGI
 Hmw2com SNSKGTLTQY RSSAGVNFG V..N...GMM SFNLKEGAKV NFKLKPNEJM
 62/68

751

800

Hmw3com FKSNANYAL. FNEDISVSG. .GGSVNFKLN ASSSNIQTPG VIIKSQNFNV
 Hmw4com DPKKELPIT. FNANITATGN SDSSVMFDIH A...NLTSRA AGINMDSINI
 Hmw1com NKYSSLNYAS FNGNISVSG. .GGSVDFTLI ASSSNVQTPG VVINSKYFNV
 Hmw2com NTSKPLPI.R FLANITATG. .GGSVFFDIY ANHS...GRG AELKMSEINI

801

850

Hmw3com SGGSTLNKA EGSTETAFSI ENDLNLNATG GNITIRQVEG T..DSRVNKG
 Hmw4com TGGLDFSITS HNRNSMAFEI KKDLTINATG SNFSLKQTKD SFYNEYSKHA

FIG. 10G.

Hmw1.com STGSSLRFKT SGSTKTFGFSI EKDLTINATG GNITLLQVEG T . DGMIGKG
 Hmw2.com SNGANFTLNS HVRGDDAFKI NKDLTINATN SNFSLRQTKD DFYDGYARNA

851 900
 Hmw3.com VAAKKNITFK GGNITFGSQK ATTEIKGNVT INKNNTNATLR GANFAEN . . .
 Hmw4.com INSSHNLTIL GGNVTLGGEN SSSSITGNIN ITNKAANVTILQ ADTSNSNTGL

63 / 68
 Hmw1.com IVAKKNITFE GGNITFGSRK AVTEIEGNVT INNNANVTLLI GSDFDNHQ . .
 Hmw2.com INSTYNISIL GGNVTLGGQN SSSSITGNIT IEKAANVTILE ANNAPNQQNI

901 950
 Hmw3.com KSPLNIAGNV INNGNLTTAG SIINIAGNLT VSKGANLQAI TNYTFNVAGS
 Hmw4.com KKRTLTLGNI SVEGNLSLTG ANANIVGNLS IAEDSTFKGE ASDNLNITGT
 Hmw1.com KPLTIKKDVI INSGNLTTAG NIVNIAGNLT VESNANFKAI TNFTFNVGGL
 Hmw2.com RDRVVIKLGSL LVNGSLSLTG ENADIKGNLT ISESATFKGK TRDTLNITGN

951 1000
 |||

FIG. 10H.

Hmw3.com FDNNNGASNIS IARGGAKFK. DINNTSSLNI TTNSD'TTYRT IIKGNISSNKS
 Hmw4.com FTNNNGTANIN IKQGVVKLQG DINNKGGLNI TTNASGTQKT IINGNITNEK
 Hmw1.com FDNKGNNSNIS IAKGGARFK. DIDNSKNLSD TTNSSSTYRT IISGNITTNKN
 Hmw2.com FTNNNGTAEIN ITQGVVKLG. NVTNDGDLNI TTHAKRNQRS IIIGGDIINNK

1001 1050 64 / 68
 Hmw3.com GDLNITDKKS DAEIQIGGNI SQKEGNLTIS SDKVNITNQI TIKAGVEGGR
 Hmw4.com GDLNIKNIKA DAEIQIGGNI SQKEGNLTIS SDKVNITNQI TIKAGVEGGR
 Hmw1.com GDLNITNEGS DTEMQIGGDI SQKEGNLTIS SDKINITKQI TIKAGVGDGEN
 Hmw2.com GSLNITDSNN DAEIQIGGNI SQKEGNLTIS SDKINITKQI TIKKGIDGED

1051 1100
 Hmw3.com SDSSEAENAN LTIQTKEKL AGDLNISGFN KAEITAKNGS DLTIGNASGG
 Hmw4.com SDSSEAENAN LTIQTKEKL AGDLNISGFN KAEITAKNGS DLTIGNASGG
 Hmw1.com SDSDATNNAN LTIKTKEKL TDQLNISGFN KAEITAKDGS DLTIGNNTNSA
 Hmw2.com SSSDATSNAN LTIKTKEKL TEDLSISGFN KAEITAKDGR DLTIGNSNDG

FIG.10I.

1101 1150

Hmw3com N..ADAKKVT FDKVKDSKIS TDGHNVTLNS EVKT.. SNGS SNAGNDNSTG
 Hmw4com N..ADAKKVT FDKVKDSKIS TDGHNVTLNS EVKT.. SNGS SNAGNDNSTG
 Hmw1com D.GTNAKKVT FNQVKDSKIS ADGHKVTLHS KVETSGSNNN TEDSSDMNAG
 Hmw2com NSGAEEAKKVT FNNVKDSKIS ADGHNVTLNS KVKTSSSNNGG RESNSNDTG

1151 1200 65 / 68

Hmw3com LTISAKDVTV NNNVTSHKTI NISAAAGNVT TKEGTTINAT TGSVEVTAQN
 Hmw4com LTISAKDVTV NNNVTSHKTI NISAAAGNVT TKEGTTINAT TGSVEVTAQN
 Hmw1com LTIDAKNVTV NNNITSHKAV SISATSGEIT TKTGTTINAT TGNVEIT...
 Hmw2com LTITAKNVEV NKDVTSLKTV NITA. SEKVT TTAGSTINAT NGKASIT...

1201 1250

Hmw3com GTIKGNITSQ NVTVTATENL VTTENAVINA TSGTVNISTK TGDIKGGIES
 Hmw4com GTIKGNITSQ NVTVTATENL VTTENAVINA TSGTVNISTK TGDIKGGIES
 Hmw1comAQ TGDIKGGIES

FIG. 10J.

Hmw2.com TK T

1251	Hmw3.com	TSGNVNITAS	GNTLKVSNIT	QQDVTVTADA	GALT TAGST	ISATTGNANI
	Hmw4.com	TSGNVNITAS	GNTLKVSNIT	QQDVTVTADA	GALT TAGST	ISATTGNANI
	Hmw1.com	SSGSVTLTAT	EGALAVSNIS	GNTVTVTANS	GALT LAGST	IKG.TESVTT
	Hmw2.com

66 / 68

1350

Hmw3.com TTKTGADINGK VESSSSGSVTL VATGATLAVG NISGNTVTIT ADSGKLITSTV
Hmw4.com TTKTGADINGK VESSSSGSVTL VATGATLAVG NISGNTVTIT ADSGKLITSTV
Hmw1.com SSQSGDIG. G TISGGTVEVK ATESLTQSN
Hmw2.com GDIS. G TISGNTVSVS ATVDLTTKSG

1301

Hmw3.com TTKTGADINGK VESSSSGSVTL VATGATLAVG NISGNTVTIT ADSGKLITSTV
Hmw4.com TTKTGADINGK VESSSSGSVTL VATGATLAVG NISGNTVTIT ADSGKLITSTV
Hmw1.com SSOSESDIG

1351	Hmw3.com	GSTINGTNSV	TTSSQSGDIE	GTISGNtvnv	TASTGDLTIG	NSAKVEAKNG
	Hmw4.com	GSTINGTNSV	TTSSQSGDIE	GTISGNtvnv	TASTGDLTIG	NSAKVEAKNG

FIG. 10K.

Hmw1 com SKIKATTGEA NVTSATGTIG GTISGNTVNV TANAGDLTVG NGAEIFATEG
 Hmw2 com SKIEAKSGEA NVTSATGTIG GTISGNTVNV TANAGDLTVG NGAEIFATEG

1401

1450
 Hmw3 com AATLTAESGK LTTQTGSSIT SSNGQTTLTA KDSSIAGNIN AANVTLNITG
 Hmw4 com AATLTAESGK LTTQTGSSIT SSNGQTTLTA KDSSIAGNIN AANVTLNITG
 Hmw1 com AATLTTSSGK LTTEASSHIT SAKGQVNLSA QDSSVAGSIN AANVTLNITG
 Hmw2 com AATLTATGNT LTTEAGSSIT STKGQVDLLA QNSSIAGNIN AANVTLNITG
 67 / 68

1451

1500
 Hmw3 com TLTTTGDSKI NATSGTLTIN AKDAKLDGAA SGDRTVVNAT NASGSGNVTA
 Hmw4 com TLTTTGDSKI NATSGTLTIN AKDAKLDGAA SGDRTVVNAT NASGSGNVTA
 Hmw1 com TLTTVKGSNI NATSGTLTIN AKDAELNGAA LGNHTVVNAT NANGSGSVIA
 Hmw2 com TLTTVAGSDI KATSGTLTIN AKDAKLNDA SGDSTEVNAV NASGSGSVTA

1501

1550

FIG. 10L.

Hmw3.com	KTSSSVNITG	DLNTINGLNI	ISENGRNTVR	LRGKEIDVKY	IQPGVASVEE
Hmw4.com	KTSSSVNITG	DLNTINGLNI	ISENGRNTVR	LRGKEIDVKY	IQPGVASVEE
Hmw1.com	TTSSRVNITG	DLITINGLNI	ISKNGINTVL	LKGVKIDVKY	IQPGLIASVDE
Hmw2.com	ATSSSVNITG	DLNTVNGLNI	ISKDGRNTVR	LRGKEIEVKY	IQPGVASVEE

1551	1600				
Hmw3.com	VIEAKRVLK	VKDLSDDEERE	TLAKLGVSAV	RFVEPNNAIT	VNTQNEFTTK
Hmw4.com	VIEAKRVLK	VKDLSDDEERE	TLAKLGVSAV	RFVEPNNAIT	VNTQNEFTTK
Hmw1.com	VIEAKRILEK	VKDLSDDEERE	ALAKLGVSAV	RFIEPNNTIT	VDTQNEFATR
Hmw2.com	VIEAKRVLK	VKDLSDDEERE	TLAKLGVSAV	RFVEPNNTIT	VNTQNEFTTR

1601	1632			
Hmw3.com	PSSQVITISEG	KACFSSGNGA	RVCTNVADDG	QQ
Hmw4.com	PSSQVITISEG	KACFSSGNGA	RVCTNVADDG	QQ
Hmw1.com	PLSRIVISEG	RACFSNSDGA	TVCVNIAIDNG	R.
Hmw2.com	PSSQVITISEG	KACFSSGNGA	RVCTNVADDG	QP

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US94/02550

A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) : A61K 39/02

US CL : 424/92

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 424/92; 435/851

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Gene-Seq, APS, Biosis, Embase, Scisearch, Chem Abstracts

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Pediatric Infectious Disease Journal, Volume 9, No. 5, issued 05 May 1990, Barenkamp et al, "Development of Serum Bactericidal Activity Following Nontypable Haemophilus influenzae Acute Otitis Media", pages 333-339, see page 337.	1-3
Y	Pediatric Research, Volume 29, No. 4 part 2, issued 1991, Barenkamp S. J., "DNA Sequence Analysis of Genes for Nontypable Haemophilus influenza High Molecular Weight Outer Membrane Proteins which are Targets of Bactericidal Antibody", see page 167A, column 1, abstract no. 985.	1-3

<input type="checkbox"/>	Further documents are listed in the continuation of Box C.	<input type="checkbox"/>	See patent family annex.
•	Special categories of cited documents:	• T	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
• A	document defining the general state of the art which is not considered to be part of particular relevance	• X	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
• E	earlier document published on or after the international filing date	• Y	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
• L	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	• &	document member of the same patent family
• O	document referring to an oral disclosure, use, exhibition or other means		
• P	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search	Date of mailing of the international search report
09 MAY 1994	JUN 02 1994
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	Authorized officer JULIE KRSEK-STAPLES Telephone No. (703) 308-0196
Facsimile No. (703) 305-3230	<i>Jill Warden for</i>

Form PCT/ISA/210 (second sheet)(July 1992)*

THIS PAGE BLANK (USPTO)